Genomic Characterization of Familial Pancreatic Cancer

Leads to Discovery of a Novel Structural Variant in Cancer

by
Alexis Leigh Norris

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
March 2015

Abstract

Despite dramatic increases in the survival of cancer in general over the last 5 decades, the 5-year survival of pancreatic ductal adenocarcinoma (PDAC) remains relatively unchanged at 6%. Approximately 10% of PDAC cases are familial pancreatic cancer (FPC), involving two or more affected first-degree relatives. Despite large sequencing efforts over the past decade, less than 20% of FPC cases have an identified causal germline mutation, despite recent large-scale sequencing efforts by our group and others.

Unlike the majority of familial cancer syndromes, our analysis shows that FPC is not associated with an earlier age of onset compared to its sporadic counterpart. We theorized that the similar age of onset may be the result of shared between sporadic and familial PDAC driver genes, *KRAS*, *TP53*, *CDKN2A/p16*, and *DPC4/SMAD4*. Our analysis of PDAC driver genes in our cohort of FPC cancer cell lines, using high density SNP microarray, whole exome sequencing (WES), whole genome sequencing (WGS), and RNA-Seq, confirmed that FPC has alterations in the PDAC driver genes and at the same prevalence as sporadic.

Given that the genes *p16* and *SMAD4* are commonly inactivated by deletions, we hypothesized that the underlying structural rearrangements could provide insight into the mechanism of driver gene deletions in cancer. Our breakpoint analysis of the *p16* and *SMAD4* deletions using WGS data revealed a novel structural variant that we have termed "TransFlip mutations." A TransFlip mutation is an inter-chromosomal translocation on one side and an inversion on the other side, flanking a deletion.

My thesis work highlights the similarity of FPC with sporadic PDAC, both in the driver genes and in the age of onset. Here I report the discovery of TransFlip mutations, a new structural variant in cancer, highlighting the complexity of cancer genomes and the challenges of reliably calling structural variants.

Advisor:

Dr. James R. Eshleman (Reader)

Thesis Committee:

Dr. Sarah J. Wheelan (Reader)
Dr. Kenneth W. Kinzler (Chair)
Dr. Ralph H. Hruban
Dr. Alison P. Klein

Acknowledgements

I would first like to thank my mentor Dr. James Eshleman, for having faith in my ability to be an independent scientist from day one. He has been supportive, encouraging, and inspirational throughout my time here, and allowed me the freedom to pursue research that excites me. My thesis committee members Drs. Kenneth Kinzler, Ralph Hruban, Alison Klein, and Sarah Wheelan provided critical input to my thesis project, and I appreciate all their time and efforts. I also thank my lab members and the larger G.I. Pathology group for their technical advice and support. Specifically, I would like to thank Dr. Niki Ottenhof, whose friendship and support has never wavered, even now from across the Atlantic Ocean.

I thank my undergraduate mentors Drs. Carol Ely-Hepfer and Richard Bamford for teaching me how to use a pipette and pushing me to pursue academic research. Then as a technician at Johns Hopkins, it was my mentors Drs. Kathleen Murphy and Constance Griffin who provided an opportunity for research that led to my first publications. I deeply thank both of them, as well as all of my colleagues, past and present, at the Molecular Diagnostics Lab.

Once I began the Cellular and Molecular Medicine (CMM) graduate program, Dr. Rajini Rao, Colleen Graham, and Leslie Lichter became my support team, and I am grateful for their help getting me through those first years. My CMM classmates became my new family during those long days and nights studying for our qualifying exam. A special thanks to my classmates Allie Galanis (Western Blot expert), Sean Cho (R expert), Aggie Rucki (mouse expert), and Nina Hosmane (cell culture expert), for both their scientific advice and friendship.

Thank you Dr. Benjamin Orsburn for reminding me every day how much I love science, pugs, and you. Finally, I would like to thank my family, including my grandfather Dr. Vincent Norris, for everything you have done for me throughout the years. From an early age, you fostered my scientific curiosity and urged me to pursue a career that I love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Familial Pancreatic Cancer

Pancreatic ductal Adenocarcinoma (PDAC) has a dismal 5-year survival rate of 6%, which has remained relatively unchanged for decades, despite huge improvements in the survival of other cancers (*1*). The well-defined molecular progression of PDAC includes the driver genes *KRAS*, *TP53*, *CDKN2A/p16*, and *DPC4/SMAD4* (*2*). About 10% of PDAC cases occur in familial aggregate of two or more affected first-degree relatives, the definition of familial pancreatic cancer (FPC) (*3, 4*). Early onset is a hallmark of most familial cancer syndromes, including hereditary breast and ovarian cancer (*BRCA1*, *BRCA2*), familial adenomatous polyposis (*APC*), hereditary non-polyposis colorectal cancer (*MLH1*, *MSH2*, *MSH6*, *PMS2*), and familial atypical multiple mole melanoma syndrome (*p16*) (*5-8*). In contrast, an earlier age of onset is not an obvious hallmark of FPC (*9, 10*). In **Chapter 2**, I perform a qualitative analysis of the literature to confirm the common opinion that the age of onset of familial and sporadic PDAC cases is similar.

Identifying Biomarkers for Early Detection

Early detection alone could increase the 5 year survival 5-fold, given that the 5-year survival rate of a localized tumor approached 25% (*1*). Recent research has highlighted that the window of opportunity for early detection is over a decade (*11*). Molecular testing of non-invasive liquid biopsies can make early detection practical in a broader population. Such molecular testing could utilize tumor-specific mutations in PDAC driver genes (e.g. *KRAS*) or tumor-specific structural variants (e.g. personalized analysis of rearranged ends, PARE) (*12, 13*).

In **Chapter 2**, I show that the sporadic PDAC driver genes are also altered in FPC, and at a similar prevalence. Because FPC shares the sporadic PDAC signature mutations, these could be included in early detection tests and the gene panels currently in development could also be used in FPC kindreds for early detection. The shared driver genes may also explain the similar age of onset in familial and sporadic PDAC.

PARE analysis relies on tumor-specific structural variants (SVs), which can be identified from paired-end next generation sequencing data. In **Chapter 3**, I characterize the SV landscape of FPC homozygous deletions (HDs) at nucleotide-resolution using the combination of high density SNP

microarray and paired-end whole genome sequencing (WGS); this information is critical when designing PARE assays.

Discovery of TransFlip Mutations

During our SV analysis in **Chapter 3**, I discovered a novel class of SVs that are the combination of a translocation on one side of the deletion and an inversion on the other side, which we designate as "TransFlip mutations." Strikingly, TransFlip mutations are enriched at deletions of the PDAC driver genes *p16* and *SMAD4*. TransFlip mutations are ideal PARE targets, as they have not just one, but two novel junctions and often delete TSGs, thus providing selective advantage to the tumor cell.

Challenges in the search for new FPC predisposition genes

The known FPC predisposition genes account for less than 20% of FPC cases, and include *BRCA2*, *PALB2*, *ATM*, *STK11/LKB1*, *CDKN2A/p16*, *APC*, *PRSS1*, *SPINK1*, and the mismatch repair genes (*MLH1*, *MSH2*, *MSH6*, *PMS2*) (Table 1) (*14-17*). In **Chapter 4**, I outline our strategy we used to identify new FPC predisposition genes, using an integrated approach of conventional karyotype, high density SNP microarray, whole exome sequencing (WES), whole genome sequencing (WGS), and RNA-Seq. The lack of a bona fide candidate predisposition gene, despite our extensive analysis, highlights the challenges in identifying new FPC predisposition genes, and raises the possibility that this cancer results from multiple genetic and epigenetic alterations in addition to or even instead of sequence changes in a small number of genes.

Table 1.1. Known FPC predisposition genes

Syndrome	Gene(s)	Approx. Additional PDAC Risk
Peutz-Jeghers Syndrome (PJS)	*STK11/LKB1*	132X
Hereditary Pancreatitis	*PRSS1, SPINK1*	50-90X
Familial Atypical Multiple Mole Melanomas (FAMMM)	*CDKN2A/p16*	13-22X
Hereditary Breast and Ovarian Cancer Syndrome (HBOC)	*BRCA2, PALB2*	3-10X
Hereditary Non-Polyposis Colorectal Cancer (HNPCC)	*MLH1, MSH2, MSH6, PMS2*	9X
Familial Adenomatous Polyposis (APC)	*APC*	Up to 4X
ATM-related FPC	*ATM*	unknown

Chapter 2

Familial and sporadic pancreatic cancer share the same molecular pathogenesis

Norris AL, Roberts NJ, Jones S, Wheelan SJ, Papadopoulos N, Vogelstein B, Kinzler K, Hruban RH, Klein AP, Eshleman JR. Familial and sporadic pancreatic cancer share the same molecular pathogenesis. *Fam Cancer*. (In Press).

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is nearly uniformly lethal, with a median overall survival in 2014 of only 6 months. The genetic progression of sporadic PDAC (SPC) is well established, with common somatic alterations in *KRAS*, *p16/CDKN2A*, *TP53*, and *SMAD4/DPC4*. Up to 10% of all PDAC cases occur in families with 2 or more affected first-degree relatives (familial pancreatic cancer, FPC), but these cases do not appear to present at an obviously earlier age of onset. This is unusual because most familial cancer syndrome patients present at a substantially younger age than that of corresponding sporadic cases. Here we collated the reported age of onset for FPC and SPC from the literature. We then used an integrated approach including exomic sequencing (WES), whole genome sequencing (WGS), RNA sequencing (RNA-Seq), and high density SNP microarrays to study a cohort of FPC cell lines and corresponding germline samples. We show that the four major SPC driver genes are also consistently altered in FPC and that each of the four detection strategies was able to detect the mutations in these genes, with one exception. We conclude that FPC undergoes a similar somatic molecular pathogenesis as SPC, and that the same gene targets can be used for early detection and minimal residual disease testing in FPC patients.

Introduction

While patients with several different forms of cancer survive longer after diagnosis than in the past, the 5-year survival rate of patients with pancreatic ductal adenocarcinoma cancer (PDAC) has remained relatively unchanged over the past 5 decades (*1*). As many as 10% of PDAC have a hereditary component (familial PDAC, FPC), defined as at least two first-degree relatives with PDAC (*3, 4*). Known susceptibility genes include *BRCA1*, *BRCA2*, *PALB2*, *ATM*, *STK11*, *PRSS1*, *SPINK1*, and DNA mismatch repair genes, but all together these explain less than 20% of familial pancreatic cancer cases (*14-17*). PDAC is notoriously lethal because patients present late in the disease process and the cancers are chemorefractory. Importantly, the 9% of cases that present with the tumor confined to the pancreas have a 5-year survival rate of 24%, supporting the notion that lesions detected early enough can be cured (*1*). To focus early detection resources, it is important to identify patients at particularly high risk, such as those with familial predispositions.

The molecular progression of SPC is well-established both histologically and molecularly (*2*). The high-prevalence SPC driver genes are *KRAS* (>90% of PDAC), *CDKN2A/p16* (95%), *TP53* (50-75%), and *SMAD4/DPC4* (55%) (*18-21*). PDACs commonly arise from pancreatic intraepithelial neoplasia (PanIN) or Intraductal Papillary Mucinous Neoplasm (IPMN) precursor lesions. While the pathology of FPC has been shown not to differ from that of apparently sporadic disease, FPC patients have been shown to have significantly more precursor lesions as well as higher grade precursor lesions when compared to patients with sporadic disease (*22-24*). Knowing the genes involved in FPC molecular progression is essential to designing effective early detection strategies (*25*).

Early onset is a hallmark of most familial cancer syndromes, including hereditary breast and ovarian cancer (*BRCA1*, *BRCA2*), familial adenomatous polyposis (*APC*), hereditary non-polyposis colorectal cancer (*MLH1*, *MSH2*, *MSH6*, *PMS2*), and familial atypical multiple mole melanoma syndrome (*p16*) (*5-8*). In contrast, an earlier age of onset is not an obvious hallmark of FPC (*9, 10*). How can one possibly inherit a predisposition to a cancer without an obvious acceleration of the phenotype? This question challenges our current understanding of familial cancer syndromes and the canonical two-hit hypothesis (*26, 27*).

In the study, we first collated the age of onset in FPC and SPC reported in the literature to validate the opinion that the age of onset of familial and sporadic PDAC cases was similar. We then determined the status of known SPC driver genes in our own FPC cohort, a unique resource of eighteen FPC cell lines that we have generated over the past decade. We used an integrated approach including high density SNP microarrays, exomic sequencing, whole genome sequencing, and RNA-sequencing to investigate those genes involved in FPC progression. Finally, having established a consensus for each gene in each sample, we examined the ability of each tool to detect the mutations.

Materials and Methods

Case Selection

This study was reviewed and approved by the Institutional Review Board at Johns Hopkins Medical Institutions, and informed consent was obtained from all study participants. Familial pancreatic cancer was defined as a pancreatic cancer that arose in a patient with at least a first-degree relative with pancreatic cancer (i.e. 2 or more affected first-degree relatives). Cancer cell lines were established from familial pancreatic cancers and matched normal DNA from the patients was obtained from Epstein-Barr virus (EBV) transformed lymphoblasts or frozen tissue (*28*). Tumor-normal pair matching was confirmed by STR analysis of 9 loci and Amelogenin using ABI Profiler kit (Life Technologies, Carlsbad, CA) and size-separated on an ABI CE3130xl instrument (Life Technologies). The data from 94 SPC and 7 FPC (4 from discovery, 3 from prevalence) were previously reported (*20*). An unpaired, two-tailed t-test was used to determine if the mean age of onset difference between our familial and sporadic cases was statistically significant.

Collation of Reported Age of Onset

Literature reporting age of onset in FPC (excluding hereditary pancreatitis) and SPC were collected from PubMed. Only the most recent study was used when multiple studies employed the same patient registry, on the assumption that previous reported families would be included in subsequent reports and therefore exclude redundant cases. Studies were stratified based on study type (population or referral) and statistic reported (mean or median).

Preparation of Genomic DNA and RNA

Genomic DNA was extracted from early passage cell lines and matched normal EBV-transformed lymphoblasts or frozen normal tissue using QIAamp DNA mini kit (Qiagen, Valencia, CA), per manufacturer's instruction. RNA was extracted from cell lines using RNeasy mini kit (Qiagen), per manufacturer's instruction. A HPDE (human pancreatic ductal epithelium) cell line was used as a normal control for RNA-Sequencing (*29*).

High Density SNP Microarray

The Omni2.5 array (Illumina, San Diego, CA) was used to analyze cancer cell lines and matched normal samples at 2,379,855 (2.5M) SNP loci. Analysis was carried out with Genome Studio with the following criteria: an average LogR Ratio (LRR) ≤ -2.0 for homozygous deletions; LRR of 0-0.53 and B Allele Frequency of 0 or 1 for loss of heterozygosity (LOH); and an average LRR ≥ 1.4, with at least 1 SNP LRR ≥ 2.0, for amplifications. At least 4 SNPs must fit criteria for the region to be called an alteration and boundaries were the first and last SNPs that meet criteria. Adjacent deleted or amplified regions (within 100kb) were considered to be one alteration. Given that half or more of the *p16* and *SMAD4* inactivations are homozygous deletions, we excluded the 4 FPC and 81 SPC cases without SNP microarray data, in the analysis of *p16* and *SMAD4* genes.

Genomic DNA Libraries and Exomic Sequencing

Genomic DNA libraries were prepared using 1 μg of genomic DNA and human exome capture was performed following a modified protocol from Agilent's SureSelect Paired-End Version 2.0 Human Exome Kit (Agilent, Santa Clara, CA) as previously described (*30*). Briefly, captured DNA libraries were sequenced with GAIIx Genome Analyzer, yielding 150 bp (2 × 75 bp) from the final library fragments, to 200x coverage. Sequencing reads were analyzed and aligned to human genome hg18 with the Eland algorithm in CASAVA 1.7 software (Illumina). The Database of Single Nucleotide Polymorphisms was used in the analysis of whole-exome sequencing data (dbSNP). Mutations were visually confirmed in the aligned files.

Whole Genome Sequencing

Sequencing on an Illumina HiSeq 2000 (Illumina) was carried out at 60X coverage for cancers and 30X coverage for matched normal by Personal Genome Diagnostics (Baltimore, MD) using 3 μg of genomic DNA and generating 200 bp (2 x 100 bp paired reads) per fragment. Reads were aligned to human genome (hg19) with Eland v.2 algorithm in CASAVA 1.7 software (Illumina).

cDNA Libraries and RNA Sequencing

A total of 5 μg of total RNA was depleted of ribosomal RNA using ribominus and cDNA libraries were prepared using TruSeq Stranded Total RNA Sample Preparation (Illumina), per the manufacturer's

instructions. Paired-end sequencing, resulting in 100 bp reads was carried out on an Illumina HiSeq to a level of 50M reads. RSEM was used to align the sequences to human genome hg19 (*31*). Alterations were visually confirmed using Integrated Genomics Viewer (*32*).

Results

Previous investigations have noted a similar age of onset of SPC and FPC. To comprehensively examine this, we culled studies reporting FPC and SPC age of onset and published from 1991-2013 (n=15). To avoid overweighting the same families, we used only the most recent study when multiple studies reported on the same dataset. The collated studies have reported mean or median ages of 60 to 74 for SPC patients and 52 to 69 for FPC patients (Figure 1, Supplemental Table 1). Due to potential ascertainment bias, we separated the studies that reported age of onset in a population unselected based upon family history (Figure 1A) versus those from family registries (Figure 1B). The mean age of PDAC diagnosis from 1973-2000 SEER data is 70 years (*33*). In our small cohorts of FPC and SPC, there was no obvious difference in age (FPC cohort: mean 64 years (range: 42-81); SPC cohort: 66 years (range: 36-85)) (Table 1, Supplemental Table 2). The lower age of both of the cohorts we analyzed (FPC and SPC) compared to SEER was likely attributable to ascertainment or referral bias. There appears to be a greater difference in the referral based studies, likely because the vast majority of samples in our study underwent surgical resection. We have intentionally omitted statistical comparison of the groups because of the invalidity in comparing means and medians, and the small sample size that would exist without pooling these two statistics.

The molecular progression of SPC is well-documented, with common somatic alterations in the four driver genes *KRAS*, *CDKN2A/p16*, *TP53*, and *SMAD4/DPC4*, in addition to many other low prevalence genes (*20, 21*). In an attempt to identify new FPC predisposition genes, we performed a comprehensive genomic analysis of our 16 FPC cell lines. No strong candidates for predisposition genes were identified in these samples. We also determined the mutational status of the four SPC driver genes in these 16 FPC samples as assessed by each of the four methods.

Overall, the prevalence of alterations in the four SPC driver genes was similar in the 16 FPC PDACs and the 94 SPC PDACs (Figure 2A,B). Activating *KRAS* mutations were identified in 16/16 (100%) of FPC PDACs, predominately at codon 12 (94%: 63% G12D, 19% G12V, and 13% G12R) but with one case at codon 61 (6%, Q61H) (Figure 2B, Supplemental Table 3). Of the 94 sporadic PDACs, all but one had an activating *KRAS* mutation (99%). The majority of *KRAS* mutations in the SPC PDACs were

also at codon 12 (95%: 50% G12D, 31% G12V, and 12% G12R). Four SPC PDACs had codon 61 mutations (3 Q61H, 1 Q61R), and one SPC PDAC had two different activating *KRAS* mutations (G12V and G13C).

CDKN2A/p16 was inactivated in 100% (12/12, the 4 cases without SNP microarray data were excluded) by homozygous deletion (9/12, 75%) or single base substitution with loss of heterozygosity (LOH) (3/12, 25%), of the FPC PDACs, compared to only 62% (8/13, the cases without SNP microarray data were excluded) of the SPC PDACs (p=0.04, Figure 2B, Supplemental Table 4). Alterations of the *CDKN2A* gene are reported to occur in 95% of SPC PDACs, with epigenetic silencing accounting for about 15% of this inactivation (*34*). As we did not assess epigenetic changes, the actual fraction of cases with somatically altered CDKN2A in Supplemental Table 4 is likely an underestimate.

TP53 was mutated in 88% (14/16) of FPC PDACs, by single base substitution with LOH (10/14), frameshift with LOH (2/14), or biallelic mutation (2/14) (Figure 2B, Supplemental Table 5). Of the 94 sporadic PDACs, 82 (87%) had inactivating *TP53* mutations. The mutation types included biallelic mutations (1/82), single base substitutions with LOH (64/82), frameshifts with LOH (15/82), and homozygous deletions (2/82).

SMAD4/DPC4 was inactivated in 75% (9/12, the 4 cases without SNP microarray data were excluded) of FPC PDACs, by homozygous deletion (5/9), single base substitution with LOH (3/9), and frameshift with LOH (1/9) (Figure 2B, Supplemental Table 6). Of the 13 sporadic PDACs (the cases without SNP microarray data were excluded), 62% had inactivated *SMAD4*. The mutation types included single base substitutions with LOH (2/8), frameshifts with LOH (3/8), and homozygous deletions (3/8).

Having established a consensus gene mutation status, we retrospectively determined the ability of each genome-wide tool to detect the mutations. We first categorized the mutations as homozygous deletions (HDs), point mutations (including single base substitutions, frameshift deletions and insertions), and loss-of-heterozygosity (LOH) events. We then studied the ability of each tool to detect these three types of mutations in the 4 driver genes (Table 2).

Homozygous deletions are common in the tumor suppressors *CDKN2A/p16* and *SMAD4/DPC4* and were detected reliably by SNP microarray, WES, WGS, and RNA-Seq. For only one homozygous deletion (Figure 3A,B), the standard WGS Illumina pipeline for calling copy number alterations missed a *p16* homozygous deletion (sample PA222C), clearly deleted by visual inspection of WGS data (Figure 3C). The homozygous deletion included 17kb of the 5' end of *p16* transcript variant 4 (NM_058195), but did not result in the deletion of any DNA sequence corresponding to transcript variants 1, 3, 5 (NM_000077, NM_058197, and NM_001195132) (Figure 3A,B). The later transcript variants encode the p16(INK4) isoform, a CDK inhibitor, while transcript variant 4 encodes a structurally distinct p14(ARF) which stabilizes TP53 by sequestering MDM2. Both isoforms are normally expressed in the pancreas. Importantly, neither the p14ARF or p16INK4a transcripts are expressed according to the RNA-Seq data (Figure 3E), a result of the loss of p14ARF's first exon and p16(INK4)'s promoter sequence, respectively. This 37kb homozygous deletion was identified by WES, RNA-Seq, and high density SNP microarray, and the deletion's breakpoints were remarkably concordant across these methods (Figure 3B-F). Because the WGS results were initially discordant, we used multiplex ligation-dependent probe amplification (MLPA) to confirm the homozygous deletion (Figure 3G). That only 1 of the 4 alternative transcripts is included in the homozygous deletion explains why this was missed by the WGS using the standard Illumina pipeline. This highlights the importance of the reference transcript used in a NGS mapping algorithm and the potential utility of remapping to known deletions, such as *p16* in the case of PDAC, especially at lower read depths.

Point mutations and LOH in *KRAS*, *p16*, *TP53*, and *SMAD4* were all detected by WES, WGS, and RNA-Seq. High density SNP microarray could of course not detect any of the point mutations in the four driver genes. However, it is likely that a custom SNP microarray could be designed to detect mutations in hotspots in *KRAS*. Where LOH in *p16*, *TP53*, and *SMAD4* genes was detected, it was detected equally by all of the methods. In the 2 cases with biallelic *TP53* mutations, there was no evidence of LOH, as expected.

We also investigated the mutation status of genes implicated at a lower frequency in PDAC, but reported to be mutated in cystic precursors, pancreatic neuroendocrine tumors (PanNETs), or implicated as

FPC predisposition genes. *MLL3* has been reported to be mutated in 9% of PDACs (*20, 35*). Here, *MLL3* was mutated in 17% (2/12) FPC PDACs. Both cases were single base substitutions (nonsense mutation with LOH in PA11X and bi-allelic missense mutations in PA18C). The genes implicated in pancreatic cystic lesions (*GNAS*, *RNF43*, *CTNNB1*, and *VHL*) were not mutated in any FPC case (*36, 37*). *ATRX*, *DAXX*, and *MEN1* are reported to be mutated in PanNETs, and *ATRX* was homozygously deleted in 1 (8%) FPC case (PA102C) (*30*). *DAXX* and *MEN1* however, were not mutated in any FPC PDACs, and no clearly deleterious mutations were identified in *ATM*, *STK11*, *PRSS1*, *PALB2*, *BRCA2*, or *SPINK1*.

Discussion

We confirm, through our qualitative analysis of the literature, that most studies do not indicate a large difference in the age of onset between SPC and FPC. While some studies do show a slight difference in the age of onset, part of this difference could be due to ascertainment biases. Some studies have shown a slightly lower age of onset in FPC compared to SPC in their cohorts (*38-44*). One study even showed a slightly later onset in their FPC group (*45*), however most studies to date have shown a similar age of onset (Supplemental Table 1) (*46-55*). The field would benefit from a rigorous meta-analysis of FPC age of onset.

Our study also showed that FPC PDACs harbor the same high prevalence genetic alterations that have been identified in SPC PDACs (Figure 2B). One purpose of analyzing driver data prevalence is to identify "holes", genes with lower than expected mutation prevalence, under the hypothesis that a homologue or pathway-related gene could be defective in the germline. A similar approach led to the elegant discovery of germline MYH mutations in familial colorectal cancers that were phenotypically similar to attenuated familial adenomatous polyposis, but lacked germline APC mutations (*56*). From our data, there are no such "holes." This finding confirms and builds upon a previous study that found that familial and sporadic pancreatic cancers had similar prevalence of mutations in the three SPC driver genes they assessed (*34*).

The late detection of pancreatic cancers contributes to the lethality of the disease. Much work has been done in the area of non-invasive early detection tests, using molecular signatures of pancreatic cancer – notably, the *KRAS* codon 12/13 mutation hotspot. Because FPC shares the SPC molecular signature mutations, these could be included in early detection tests and the gene panels currently in development could also be used in familial pancreatic kindreds for early detection and molecular relapse.

Assuming that FPC and SPC have a similar age of onset, how can one inherit a predisposition to a disease without accelerating its age of onset? Unfortunately, our study did not provide any great insights into this question and it remains unanswered. We note, however, that there is precedent in PDAC, even when the causative genes are known. For example, in patients with Familial Atypical Multiple Mole Melanoma (FAMM) syndrome, *p16* germline mutations confer a significantly earlier age of onset for

melanoma, but not for PDAC (*6, 52, 57, 58*). These observations support the idea that it is the pancreatic tissue rather than the gene that is responsible for the curious lack of age dependence on the presence of hereditary predisposition genes. The mechanisms underlying this difference represent an important area for future study as it may shed light on PDAC pathogenesis in general.

We employed an integrative strategy to more comprehensively detect alterations in FPC than previous reports. Combining WES, WGS, and RNA-Seq allowed for a greater coverage of gene-coding regions, particularly in expressed genes (Table 2). The importance of gene transcript choice in identifying alterations, such as homozygous deletions, in next generation sequencing data was highlighted by the PA222C homozygous deletion of *p16* initially missed by WGS analysis, but obvious upon visual inspection of the reads (Figure 3). Other than this one example, the methods were remarkably concordant. High density SNP microarrays strengths are identifying LOH and large homozygous deletions, both hallmarks of tumor suppressor genes.

We conclude that FPC and SPC undergo similar pathogenesis permitting the same gene targets to be used for early detection and minimal residual disease testing.

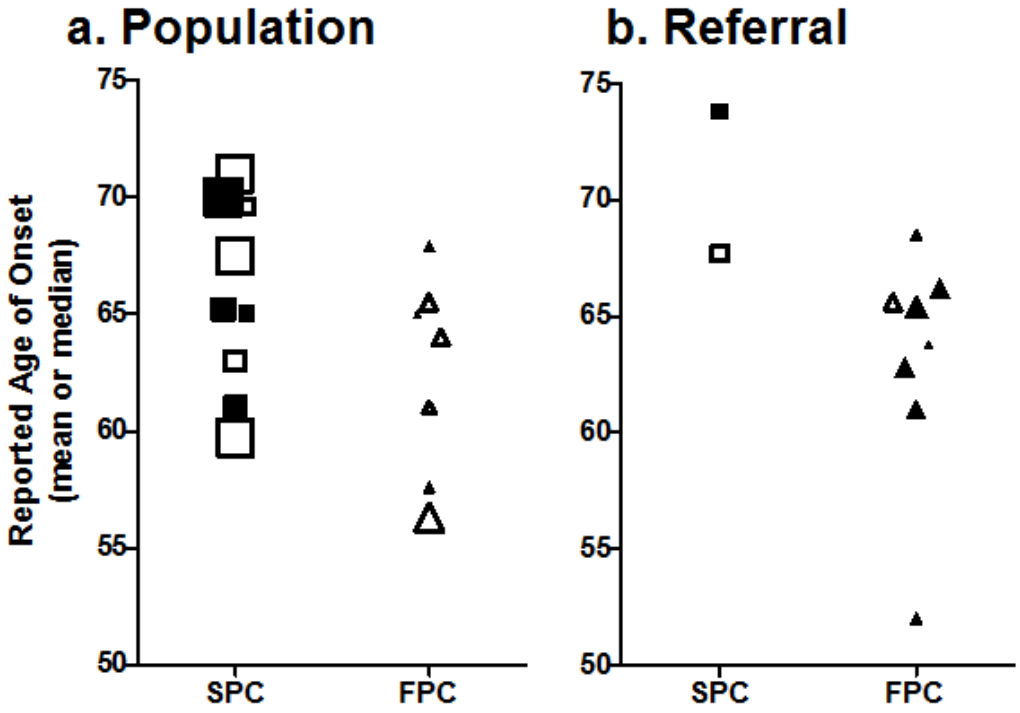
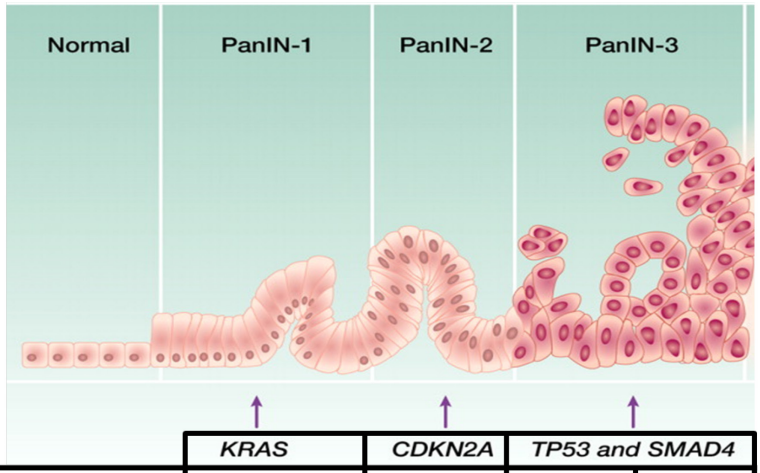


Figure 2.1. Reported age of onset for SPC and FPC, collated from the literature. Literature was separated based on **a** population-based or **b** referral cohorts, reported as means (filled symbols) or medians (empty symbols). Symbol sizes are adjusted according to the number of individuals in the study [2*log(n)]. There are no obvious differences in the age of onset for FPC (triangles) compared to SPC (squares)

a



Reported % Altered in PDAC		>90%	95%	50-75%	55%

b

Sporadic PDAC	n=94 (13)	99%	62%*	87%	62%
Familial PDAC	n=16 (12)	100%	100%*	88%	75%

Figure 2.2. Summarized alterations in PDAC molecular progression genes, for SPC and FPC. a PDAC molecular progression model with reported percent alterations of the four driver genes in PanIN lesions. **b** Percent alterations of molecular progression genes in PDAC cancers from SPC and FPC cohorts. As expected, the mutation prevalence in PDACs in panel **b** are higher than the early PanIN lesions in panel **a**. **CDKN2A*, $p = 0.04$

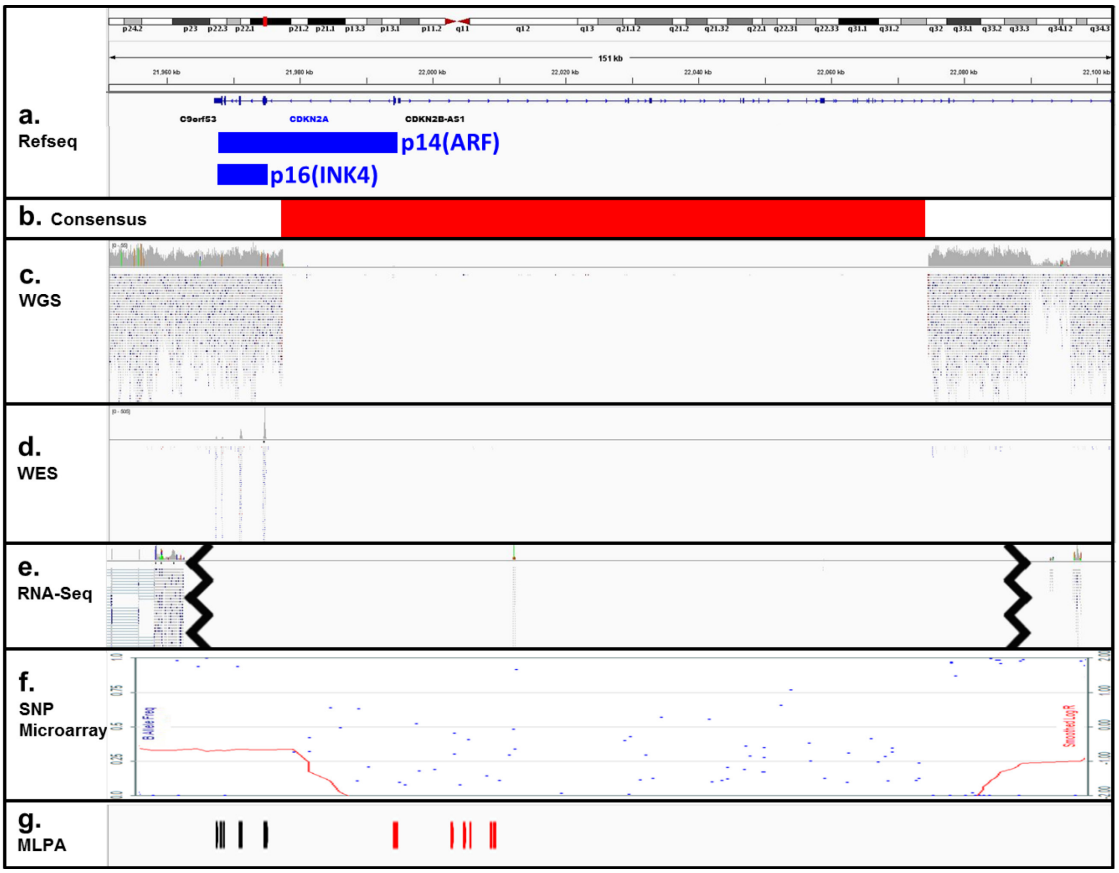


Figure 2.3. *CDKN2A/p16* homozygous deletion (HD) initially missed by WGS. Visualization of WGS reads for *p16* region (hg19, chr9:21,951,176-22,102,475) in PA222C sample using IGV and Karyostudio. The 2 protein isoforms of p16 (p14(ARF) and p16(INK4), blue) are shown as well as the adjacent genes, C9orf53 and CDKN2B-AS1 (**a**). There is clear agreement across the methods for the consensus 97kb HD boundaries (**b**). The homozygous deletion of p16 was not called by the standard Illumina pipeline for WGS data, despite a clear confirmation of the 5' HD by visual inspection, due to reference transcript choice (**c**). The HD was detected by WES, as evidenced by the lack of reads (**d**). RNA-Seq produced no high quality reads that mapped to this deleted region, but there were reads upstream (MTAP) and downstream (DMRTA1) of the HD (**e**, line break indicates upstream or downstream reads shows). High density SNP microarray detected the HD (red LogR line drops to -2.00 and scattered B allele frequencies) and the flanking LOH regions (red LogR line at -1 and B allele frequencies at 0 or 1) (**f**). MLPA probes were used to confirm the upstream LOH (black) and HD (red) regions (**g**)

Table 2.1. Familial (FPC) cohort demographics

Case	Age at Diagnosis	Sex	Additional Family History of Pancreatic Cancer
Pa007C*	80s	F	Father, Mother
Pa009C*	60s	F	Brother, Paternal Uncle
Pa011X*	70s	F	Father
Pa018X*	70s	F	Brother
Pa101C	40s	M	Father
Pa102C	60s	F	Sister
Pa147X*	70s	M	Brother, Mother
Pa170X*	40s	M	Brother
Pa212X*	60s	M	Mother
Pa222C	70s	M	Father, Paternal Grandfather, Paternal Uncle
Pa223C	60s	F	Brother, Cousin
Pa227C	60s	F	Father
Pa228C	60s	F	Brother
Pa229C	40s	M	Mother
Pa230C	60s	F	Father
Pa231C	60s	M	Father, Mother, Paternal Grandfather

*Samples previously reported in Jones, et al. Science 2008; sample set includes both cell lines (C) and xenografts (X)

Table 2.2: Relative power of each method to detect common alterations

Gene	Gene Type	Alteration Type	# Samples Altered	WES	WGS	SNP microarray	RNA-Seq
KRAS	OG	Mut	7	7 (100%)	7 (100%)	0 (0%)	7 (100%)
TP53	TSG	Mut	7	7 (100%)	7 (100%)	0 (0%)	7 (100%)
		HD	0	na	na	na	na
CDKN2A/ p16	TSG	Mut	3	3 (100%)	3 (100%)	0 (0%)	3 (100%)
		HD	4	4 (100%)	4* (100%)	4 (100%)	4 (100%)
SMAD4/ DPC4	TSG	Mut	2	2 (100%)	2 (100%)	0 (0%)	2 (100%)
		HD	4	4 (100%)	4 (100%)	4 (100%)	4 (100%)

*One *p16* mutation was not reported altered by the standard Illumina pipeline, but visualization of the bam file clearly showed a HD. Oncogene (OG), tumor suppressor gene (TSG), mutation (MUT, includes frameshifting insertions and deletions and point mutations), homozygous deletion (HD), whole exome sequencing (WES), whole genome sequencing (WGS)

Supplemental Table 2.1: Summary of the literature collated for age of onset analysis

Study	PDAC Type	Study Type	Statistic	n	Age of Onset (yrs)
James, et al. Cancer 2004	Familial	Population	Mean	30	57.6
Ghadirian, et al. Int J Pancreatol 1991	Familial	Population	Mean	14	65.0
Matsubayashi, et al. Pancreas 2011	Familial	Population	Mean	40	67.9
Brandt, et al. Ann Oncol 2008*	Familial	Population	Median	2,500	56.3
Bartsch, et al. Int J Cancer 2004	Familial	Population	Median	17	61.0
Ji, et al. Pancreatology 2008	Familial	Population	Median	150	64.0
McWilliams, et al. Clin Gastroenterol Hepatol 2006	Familial	Population	Median	74	65.5
Verna, et al. Clin Cancer Res 2010	Familial	Referral	Mean	51	52.0
Rulyak, et al. Gastroenterol 2003	Familial	Referral	Mean	176	61.0
Couch, et al. Cancer Epidemiol Biomarkers Prev 2007	Familial	Referral	Mean	350	62.8
Iqbal, et al. Br J Cancer 2012	Familial	Referral	Mean	8	63.8
Petersen, et al. Cancer Epidemiol Biomark Prev 2006	Familial	Referral	Mean	798	65.4
Mocci, et al. Cancer Epidemiol Biomarkers Prev 2013	Familial	Referral	Mean	417	66.2
Brune, et al. JNCI 2010	Familial	Referral	Mean	29	68.5
Anderson, et al. Am J Gastroenterol 2012	Familial	Referral	Median	28	65.6
James, et al. Cancer 2004	Sporadic	Population	Mean	796	61.0
Ghadirian, et al. Int J Pancreatol 1991	Sporadic	Population	Mean	165	65.0
Matsubayashi, et al. Pancreas 2011	Sporadic	Population	Mean	537	65.2
SEER_1973-2000	Sporadic	Population	Mean	72,700	70.0
Brandt, et al. Ann Oncol 2008*	Sporadic	Population	Median	25,000	59.7
Bartsch, et al. Int J Cancer 2004	Sporadic	Population	Median	456	63.0
Ji, et al. Pancreatology 2008	Sporadic	Population	Median	19,336	67.5
McWilliams, et al. Clin Gastroenterol Hepatol 2006	Sporadic	Population	Median	134	69.6
SEER18_2006-2010	Sporadic	Population	Median	40,676	71.0
Rulyak, et al. Gastroenterol 2003	Sporadic	Referral	Mean	83	69.0
Brune, et al. JNCI 2010	Sporadic	Referral	Mean	8	73.8
Anderson, et al. Am J Gastroenterol 2012	Sporadic	Referral	Median	783	67.7

*estimated

Supplemental Table 2.2: Age of onset in our SPC and FPC cohorts

Statistic	SPC	FPC
n	94	16
Min age	36	42
Max age	85	81
Mean age	65.5	64.3
SD	10.2	11.2
Median age	66	66
p-value	**0.67**	

Supplemental Table 2.3: *KRAS* mutation spectrum in FPC vs. SPC, by amino acid substitution

Alteration	FPC		SPC	
WT	0	0%	0	0%
G12D	8	67%	9	69%
G12R	0	0%	0	0%
G12V	3	25%	3	23%
Q61H	1	8%	1	8%
TOTAL	12		13	

wildtype(WT)

Supplemental Table 2.4: *CDKN2A/p16* mutation spectrum in FPC vs. SPC

Alteration	FPC		SPC	
WT	0	0%	5	38%
Biallelic mutations	0	0%	0	0%
SBS	0	0%	0	0%
Fs/indel	0	0%	0	0%
Mutation+LOH	3	25%	2	15%
SBS	2	17%	2	15%
Fs/indel	1	8%	0	0%
HD	9	75%	6	46%
TOTAL	12		13	

wildtype (WT), single base subsititution (SBS), frameshifting insertion or deletion (FS/indel), loss of heterozygosity (LOH), homozygous deletion (HD)

Supplemental Table 2.5: *TP53* mutation spectrum in FPC vs. SPC

Alteration	FPC		SPC	
WT	1	8%	0	0%
Biallelic mutations	0	0%	0	0%
SBS	0	0%	0	0%
Fs/indel	0	0%	0	0%
Mutation+LOH	11	92%	11	85%
SBS	8	67%	9	69%
Fs/indel	3	25%	2	15%
HD	0	0%	2	15%
TOTAL	12		13	

wildtype (WT), single base subsititution (SBS), frameshifting insertion or deletion (FS/indel), loss of heterozygosity (LOH), homozygous deletion (HD)

Supplemental Table 2.6: *SMAD4/DPC4* mutation spectrum in FPC vs. SPC

Alteration	FPC		SPC	
WT	3	25%	5	38%
Biallelic mutations	0	0%	0	0%
SBS	0	0%	0	0%
Fs/indel	0	0%	0	0%
Mutation+LOH	4	33%	5	38%
SBS	3	25%	2	15%
Fs/indel	1	8%	3	23%
HD	5	42%	3	23%
TOTAL	12		13	

wildtype (WT), single base subsititution (SBS), frameshifting insertion or deletion (FS/indel), loss of heterozygosity (LOH), homozygous deletion (HD)

Chapter 3

Discovery of TransFlip mutations of *p16* and *SMAD4* tumor suppressor genes

Norris AL, Kamiyama H, Makohon-Moore A, Pallavajjala A, Morsberger LA, Lee K, Batista DA, Iacobuzio-Donahue CA, Tseh-Lin M, Klein AP, Hruban RH, Wheelan SJ, Eshleman JR. TransFlip mutations produce deletions in cancer. *Genes Chromosomes Cancer*. (Revision Submitted).

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is driven by the inactivation of the tumor suppressor genes (TSGs), *CDKN2A/p16* and *DPC4/SMAD4*, commonly by homozygous deletions (HDs). Using a combination of high density SNP microarray and whole genome sequencing (WGS), we fine-mapped novel breakpoints surrounding deletions of *p16* and *SMAD4* and characterized them by their underlying structural variants (SVs). Only one third of *p16* and *SMAD4* deletions (6 of 18) were simple interstitial deletions, rather, the majority of deletions were caused by complex rearrangements, specifically, a translocation on one side of the TSG in combination with an inversion on the other side. We designate these as "TransFlip" mutations. Characteristics of TransFlip mutations are: (1) a propensity to target the TSGs *p16* and *SMAD4* ($p<0.005$), (2) not present in the germline of the examined samples, (3) non-recurrent breakpoints, (4) relatively small (47bp to 3.4kb) inversions, (5) inversions can be either telomeric or centromeric to the TSG, and (6) non-reciprocal, and non-recurrent translocations. TransFlip mutations are novel complex genomic rearrangements with unique breakpoint signatures in pancreatic cancer. We hypothesize that they are a common but poorly understood mechanism of TSG inactivation in human cancer.

Introduction

Homozygous deletions (HD) are a common mechanism by which tumor suppressor genes (TSGs) are inactivated in cancer, and are a common mapping tool for the discovery of TSGs (Hahn, et al. 1996; Lee, et al. 1987). HDs can range in size from kilobases to megabases, and are enriched at fragile sites and have been associated with chromothripsis (Bignell, et al. 2010; Stephens, et al. 2011). HDs are often the result of double-stranded DNA breaks (DSB), which can be repaired through a variety of mechanisms including homologous, microhomologous, and non-homologous pathways (Supplemental Table1). Non-homologous end joining (NHEJ), microhomology-mediated end-joining (MMEJ), and microhomology-mediated break-induced replication (MMBIR, also known as fork stalling and template switching, FoSTeS) are the dominant pathways by which DSB can be repaired, and each of these mechanisms manifests differently in the repaired DNA (Chiang, et al. 2012). Recent research has highlighted the additional mechanisms of chromoplexy, LINE1 (L1) 3' transduction, and breakage-fusion-bridge cycles (BFB) (Baca, et al. 2013; Campbell, et al. 2010; Tubio, et al. 2014).

Cyclin-Dependent Kinase Inhibitor 2A (*CDKN2A*/*p16*) and Deleted in Pancreatic Carcinoma (*DPC4*/*SMAD4*) are two of the most commonly homozygously deleted TSGs in cancer (Cox, et al. 2005). In pancreatic ductal adenocarcinoma (PDAC), *p16* and *SMAD4* homozygous deletions are seen in about 40% and 30% of cases, respectively (Biankin, et al. 2012; Caldas, et al. 1994; Hahn, et al. 1996; Jones, et al. 2008). The *p16* locus, at 9p21, encodes 2 distinct proteins, p16(INK4A) and p14(ARF1), which regulate cyclin-dependent kinases and the tumor suppressor p53, respectively. The *SMAD4* locus, at 18q21.1, encodes a transcription factor in the TGF-beta signaling pathway.

In this study, we first identified somatic HDs in PDAC cancer cell lines using high density SNP microarrays. We then determined the precise breakpoints and characterized the underlying complex structural rearrangements (SV; interstitial deletion, translocation, inversion) using paired-end whole genome sequencing (WGS). We further characterized all deletions (heterozygous and homozygous) involving the PDAC TSGs, *p16* and *SMAD4*, and confirmed rearrangements by PCR amplification and bidirectional Sanger sequencing across the breakpoints. We report two distinct patterns of TSG deletions: the first is simple interstitial deletion, and second, are deletions that result from the combination of an inter-

chromosomal translocation and inversion flanking the HD of the TSG, designated here as TransFlip mutations. For somatic HDs in other regions of the genome, TransFlip mutations are rare but are an important mechanism by which TSGs are inactivated in PDAC, where TransFlip mutations occurred more frequently than simple interstitial deletions.

Materials and Methods

Preparation of Genomic DNA and Karyotyping

Cancer cell lines were generated and matched normal DNA were obtained from 8 patients with familial pancreatic adenocarcinoma (FPC, defined as a patient with PDAC who is from a kindred in which at least two family members were diagnosed with PDAC) as previously described (Kamiyama, et al. 2013; Norris, et al. 2014). STR analysis using ABI Profiler kit (Life Technologies, Carlsbad, CA) and 3130xL Genetic Analyzer (Life Technologies) confirmed tumor-normal pair matching, and xenografting in nude mice confirmed that the cell line isolated was the neoplastic component. Genomic DNA was extracted from early passage (<20 passages) cell lines and matched normal EBV-transformed lymphoblasts or frozen normal tissue using QIAamp DNA mini kit (Qiagen, Valencia, CA), per manufacturer's instruction. Cell lines were karyotyped using methodology previously described (Griffin, et al. 1994). For comparison, DNA from the tumors of sporadic PDAC patients was obtained, as part of the rapid autopsy program previously reported (Embuscado, et al. 2005).

High Density SNP Microarray

The Omni2.5 array (Illumina, San Diego, CA) was used to analyze cancer cell lines and matched normal samples, as previously described (Norris, et al. 2014). GenomeStudio and KaryoStudio (Illumina) were used to identify regions of homozygous deletion, heterozygous deletion, and loss of heterozygosity (LOH), and matched normal was used to remove germline deletions. These approximate coordinates were used to inform breakpoint analysis of WGS data.

Whole Genome Sequencing (WGS)

Sequencing on an Illumina HiSeq 2000 (Illumina) was carried out at 60X coverage by Personal Genome Diagnostics (PGDx, Baltimore, MD) using 3µg of genomic DNA and generating 200 bp (2 x 100bp reads) sequence per fragment (median total size approximately 500bp). Paired reads were aligned to human genome (hg19) with Eland v.2 algorithm using the CASAVA 1.7 software (Illumina).

Breakpoint Characterization and Sanger Sequencing Confirmation

The exact genomic coordinates of *p16* and *SMAD4* deletion breakpoints were determined by visual inspection of aligned WGS reads in IGV (Broad Institute, Boston, MA). The approximate genomic

coordinates from SNP microarray data were used as a starting point. The reads were visually scanned in IGV for evidence of a deletion breakpoint, specifically: (1) drop in coverage, (2) mismatched bases at read ends, (3) reads with aberrant insert size, (4) reads with aberrant pair orientation, (5) reads with pair mapped to another chromosome, or (6) flanking reads with unmapped pairs. Tumor DNA and matched normal were amplified with M13-tagged primers (Integrated DNA Technologies, Coralville, IA), designed with Primer3 (Untergasser, et al. 2007), using Platinum PCR Super Mix (Life Technologies), per manufacturer's protocol. Amplicons were size-separated using polyacrylamide gel electrophoresis (PAGE) and ethidium bromide staining, to confirm the presence of novel junctions in the tumor (cancer cell line and available primary tumor DNA for Pa227C, Pa228C, Pa229C, and Pa230C) and the absence in the normal DNA. Amplicons from tumor DNA were Sanger sequenced on the 3730xL DNA Analyzer (Life Technologies) and analyzed with Sequencher software (Gene Codes, Ann Arbor, MI). To eliminate the possibility of the TransFlip mutations being specific to familial PDAC, we confirmed the presence of both an interstitial deletion and a TransFlip mutation in *p16* deletions in a set of 4 sporadic PDAC tumors' WGS data.

Results

Nucleotide Resolution of HDs

During our attempts to identify FPC predisposition genes, we closely examined HD regions identified by high density SNP microarrays in a series of well-characterized FPC cell lines (Norris, et al. 2014). Visual inspection of WGS data in IGV revealed abrupt loss of coverage at both breakpoints corresponding to regions first identified by the SNP microarray. Aberrantly mapped pairs and reads with "rainbow ends" (where half of the read IGV perfectly matches the reference and the other half has highlighted mispaired bases) flank both breakpoints (Figure 1). As expected, the mispaired bases of the "rainbow end" map to the other side of the HD for a simple interstitial deletion, the definition of split reads. Aberrantly mapped read pairs are colored by the IGV software to indicate a deletion, insertion, inversion, or translocation (Supplemental Figure 1).

Simple Interstitial Deletions Underlie Most Somatic HDs

Using this approach, we identified a total of 85 somatic HDs in our PDAC cancer cell lines from 8 FPC patients. The number of somatic HDs varied by sample, from 3 (Pa101C) to 18 (Pa229C and Pa231C), with a median of 9.5.The majority (69/85, 81%) of all somatic HDs were simple, interstitial deletions, in which a segment of a chromosome arm was deleted and the ends joined (Figure 2A, Supplemental Tables 2 and 3). The HDs associated with complex rearrangements (16/85, 19%) were categorized as inversion at one breakpoint and translocation at the other (TransFlip mutations, 7/16, 44%), inversion at both breakpoints (INV+INV, 6/16, 38%), intra-chromosomal translocation at both breakpoints (ICT+ICT 1/16, 6%), or inter-chromosomal translocation at both breakpoints (TRANS+TRANS, 2/16, 13%). Deletions associated with complex rearrangements were significantly larger than interstitial HDs, with medians of 153kb and 9kb deleted chromosome material respectively ($p<0.05$ by unpaired, two-tailed t-test) (Supplemental Table 2). Of the complex rearrangements, inversions produced the smallest HDs (median of 57kb), while SVs involving translocations (TRANS+TRANS, ICT+ICT, and TransFlip mutations) produced significantly larger HDs ($p<0.05$ by unpaired, two-tailed t-test).

While most interstitial deletions were intergenic (44/69, 64%), most (12/16, 75%) complex rearrangements resulted in the deletion of one or multiple gene(s) (Supplemental Table 3). Specifically,

TransFlip mutations were enriched at the PDAC TSGs *p16* and *SMAD4* HDs, compared to interstitials (4 of 7 vs. 5 of 69, $p<0.005$ by Fisher's exact test).

Complex TransFlip mutations underlie many *p16* and *SMAD4* deletions in pancreatic cancer

We then performed a comprehensive analysis of all *p16* and *SMAD4* deletions, including both homozygous and heterozygous deletions, in our PDAC cancer cell lines from 8 FPC patients. Deletions of *p16* and *SMAD4* were present in 7 samples each, with 18 total deletions producing the 2 heterozygous *p16* deletions (Pa102C, Pa230C), 5 homozygous *p16* deletions (Pa101C, Pa222C, Pa228C, Pa229C, Pa231C), 3 heterozygous *SMAD4* deletions (Pa101C, Pa228C, Pa230C), and 4 homozygous *SMAD4* deletions (Pa102C, Pa227C, Pa229C, Pa231C). These generated a total of 26 novel junctions (where interstitial deletions produce one novel junction and TransFlip mutations produce at least two novel junctions). All 26 novel junctions were confirmed to be present in the cancer cell line by PCR and bidirectional Sanger sequencing, and absent in the matched normal sample. For 4 cases, the resected primary tumor was available (Pa227C, Pa228C, Pa229C, and Pa230C), and all corresponding novel junctions were present, confirming that these TransFlip mutations are not an artifact of cell culture.

Complex SVs were found to underlie the majority (12/18, 66%) of *p16* and *SMAD4* deletions (Figure 2B). Most samples (5/8, 63%) had *p16* and/or *SMAD4* inactivated by a TransFlip mutation (Table 1, Table 2, Supplemental Table 4). TransFlip mutations were coupled with LOH (*p16*, Pa229C and Pa231C), with an interstitial deletion (*p16*, Pa222C), or with another TransFlip mutation (*SMAD4*, Pa102C and Pa227C), to produce a HD of the TSG. TransFlip mutations and inactivation of the other allele of the TSGs by point mutation were mutually exclusive. In 5 samples with a point mutation in combination with a deletion of the TSG, 4 were the result of an INV rearrangement centromeric to the TSG and a terminal deletion (*p16*: 1/2; *SMAD4*: 3/3). The remaining one *p16* heterozygous deletion was a simple interstitial deletion (Pa102C).

Characteristics of TransFlip mutations in pancreatic cancer

All translocations were non-reciprocal and conventional karyotyping only detected one of the translocation junctions (1/11, 9%, Pa231C's t(4;9)(p15.1;p21.3)) (Supplemental Table 5). Inversion sizes ranged from 47bp to 3.3Mb for *p16*, and 53bp to 35.3Mb for *SMAD4*, and inversions associated with

TransFlip mutations were the shortest (47bp to 3.4kb). The inversion half of a TransFlip mutation was either centromeric or telomeric to the TSG deletion (Figure 3). In addition to the characteristics above, TransFlip mutations were (1) confirmed in the primary tumors to exclude the possibility of cell culture artifact, and (2) were not present in any of the germline HDs in our samples. Distinct TransFlip mutations were also identified in sporadic PDAC samples. The fundamental features of TransFlip mutations are summarized in Table 3.

Lack of alternate mechanisms of *p16* and *SMAD4* deletion

For the simple interstitial deletions and the complex rearrangements targeting *p16* and *SMAD4*, only 1 case had non-templated base insertion at the junction, which argues against NHEJ as the dominant mechanism (Supplemental Table 6). Microhomology was seen at the breakpoint in 12 of 14 (86%) *p16* novel junctions and 7/12 (58%) *SMAD4* novel junctions. However, the median microhomology was only 1 base for both *p16* and *SMAD4* (range of 0-7 and 0-3 bases, respectively). Therefore MMEJ and FoSteS/MMBIR are unlikely to be a dominant mechanism of the *p16* and *SMAD4* rearrangements in our cohort. Only one inversion of a TransFlip mutation (Pa222C *SMAD4*) had a relative copy number increase, indicative of a fold-back inversion, a manifestation of BFB cycles (Figure 4). BFB cycles do not explain the majority of TransFlip mutations.

Chromothripsis of chromosomes 9 and 18 was seen in 1 and 2 cases, respectively (Supplemental Figure 2). Pa102C exhibited chromothripsis for both chromosomes and Pa227C exhibited chromothripsis for chromosome 18 (it did not have a *p16* deletion). Chromothripsis-associated genome remodeling has been shown to be mostly intra-chromosomal translocations (rearrangements within the same chromosome), compared to the exclusively inter-chromosomal translocations of TransFlip mutations (Stephens, et al. 2011). HDs have been associated with fragile sites, but fragile sites do not account for the spectrum of *p16* and *SMAD4* rearrangements, since the breakpoints ranged over megabases and the translocation partners were not always in fragile sites themselves (Bignell, et al. 2010).

In most cases (*p16*: 13/14, 93% and *SMAD4*: 9/12, 75%), the regions involved in the rearrangement are not actively transcribed regions (here defined as having genes within 10kb of both sides of the breakpoint), and so chromoplexy is not a dominant mechanism. Only one case (*p16* interstitial HD in

Pa101C) had repetitive elements of the same family (in this case, LINE-1) flanking both ends of the breakpoint (Supplemental Table 6). A recent report by Startek and colleagues highlights the requirement of an absolute minimum of 96% homology between LINE elements for mediating NAHR (Startek, et al. 2015). The LINE-1 elements flanking the interstitial deletion of p16 in Pa101C (L1PA4, chr9:21,297,268-21,300,617 and L1MA5A, chr9:22,121,352-22,123,446), have only 72% homology, arguing against a LINE-LINE-mediated NAHR event. L1 3' transduction was not obvious at any *p16* or *SMAD4* rearrangements, since all deletion junctions lacked the signatures of a non-templated polyA insertion and TT/AAA ORF2p endonuclease recognition sequence.

Discussion

TransFlip mutations are more prevalent in *p16* and *SMAD4* deletions than simple interstitial deletions, in contrast to deletions elsewhere where the opposite is the case. The rearrangement breakpoints in TransFlip mutations are often in gene-free regions (they do not produce fusion transcripts), but TransFlip mutations always result in the deletion of one or more genes, with enrichment at TSGs. All translocations in TransFlip mutations were non-reciprocal, which is typical of epithelial tumors (Mitelman F 2015). The inversions associated with TransFlip mutations could be either centromeric or telomeric to the TSG, ranged from 47bp to 3.4kb in length, and only one was a fold-back inversion. Fold-back inversions have been reported in PDAC, but not in combination with translocations (Campbell, et al. 2010).

To the best of our knowledge, TransFlip mutations have not been described in a somatic disease setting, but the combination of an inversion and translocation at a deletion breakpoint has been reported in recurrent translocations of renal cell carcinoma cell lines (Ali, et al. 2013). Unlike TransFlip mutations, the inversions described in these lines were centromeric and the genetic material distal to the breakpoint was deleted on the derivative chromosome (in TransFlip mutations, the distal sequence participates in a separate translocation event). It is of interest to note that there are a couple of reports of translocations associated with TSG HDs, but these studies were done at the cytogenetic level and did not report inversions (Herholz, et al. 2007; Misawa, et al. 2004). It is possible that these represent unidentified TransFlip mutations that would have been recognized if the investigators had more detailed information, such as that provided by WGS.

Previous analyses of *p16* deletion breakpoints found that the likely mechanism of repair was tissue-specific. Illegitimate V(D)J recombination was implicated in the majority of *p16* deletions in lymphoid leukemia, while evidence of NHEJ and MMEJ were seen in lung and other non-lymphoid cancers (Cayuela, et al. 1997; Raschke, et al. 2005; Sasaki, et al. 2003). While microhomology was present at most breakpoints of both TransFlip mutations and interstitial deletions, the significance of the few bases involved seems insufficient to explain the mechanism, given the expected microhomology length of 1.35 bases genome-wide by chance alone. A microhomology length of less than 7 bases is statistically insignificant, based on a geometric distribution and accounting for GC content. Only one junction of a

TransFlip mutation (the inversion half of the *p16* TransFlip mutation in Pa231C) had 7 bases of microhomology at the junction, but the corresponding translocation half of the TransFlip mutation had only 1 base of microhomology. No one mechanism (NHEJ, MMEJ, chromothripsis, chromoplexy, LINE-LINE-mediated NAHR, L1 3' transduction, or BFB cycles) explains the prevalence of complex rearrangements, specifically TransFlip mutations, as a mechanism of *p16* and *SMAD4* inactivation in our cohort.

We propose that TransFlip mutations occur by a novel mechanism, given their unique signature (Table 3). It is possible that the translocation and inversion occur simultaneously to resolve breaks in the DNA strand, or occur sequentially (Figure 5). It remains to be determined if TransFlip mutations are the result of an inversion that requires a telomere (possibly via translocation) to survive (or vice versa), or if TransFlip mutations are a manifestation of a specific DNA repair pathway defect currently unknown. What is clear is that TransFlip mutations are associated with the inactivation of TSGs, not just HDs.

It is possible that TransFlip mutations are unique to pancreatic cancer. In 2007, in their seminal work, Griffin and colleagues revealed the particularly high level of genomic instability in pancreatic cancer cell lines, which was also evident by the complex karyotypes of these distinct pancreatic cancer cell lines (Supplemental Table 5) (Griffin, et al. 2007). Genomic instability is a main driver of complex rearrangements in cancer, and often leads to the inactivation of tumor suppressor genes, including *p16* and *SMAD4*, which were also recurrently lost in the cell lines reported by Griffin in 2007. Future studies should explore the prevalence of TransFlip mutations in other cancers, to determine if TransFlips are a manifestation of a pancreatic cancer specific type of genomic instability.

Given the lack of TransFlip mutations associated with germline HDs in our cohort, we are keenly interested in whether TransFlip mutations underlie any human genetic disease, or whether they are completely unique to cancer. TransFlip mutations could be the result of an endogenous process possessed by all cells that are selected for in cancer because of TSG deletions. Alternatively, TransFlip mutations could manifest in cancer because of a fundamental defect in cancer cells, such as DNA repair, cell cycle regulation, or chromatin remodeling, or even the manifestation of the reaction of cells to a specific mutagen.

Complex rearrangements like TransFlip mutations will impact PARE (personalized analysis of rearranged ends) analysis for early detection and minimal residual disease testing (Leary, et al. 2010). For monitoring minimal residual disease in a patient using PARE, it is imperative to use tumor-specific rearrangements that are under positive selection. TransFlip mutations are ideal PARE molecules since they often delete TSGs, thus providing selective advantage to the tumor cell, and have two novel junctions that can be assessed in the testing. One challenge of TransFlip mutations and other complex rearrangement is reliably detecting them from WGS data, and it remains to be seen if SV calling algorithms detect them as well as they do interstitial deletions, especially in primary tumors where DNA from stromal cells may obscure the detection of structural variants. We emphasize the utility of high density SNP microarray to first identify the approximate breakpoints of deletions that aid in locating the junction in WGS data.

The most exciting implication of TransFlip mutations is that they are potentially targetable via synthetic lethality, similar to *BRCA* mutations conferring sensitivity to poly ADP ribose polymerase (PARP) inhibitors or by exploiting genes in the HD, like in the co-deletion of *MTAP* with *p16* (Farmer, et al. 2005; Hustinx, et al. 2005).

Our study underscores the high complexity of structural rearrangements that can occur and which may not be fully appreciated without the combined use of multiple techniques, including conventional cytogenetics, high density SNP microarray, and WGS. Further work is needed to elucidate the mechanisms that underlie these complex rearrangements and determine the prevalence of TransFlip mutations as a mechanism of TSG inactivation.

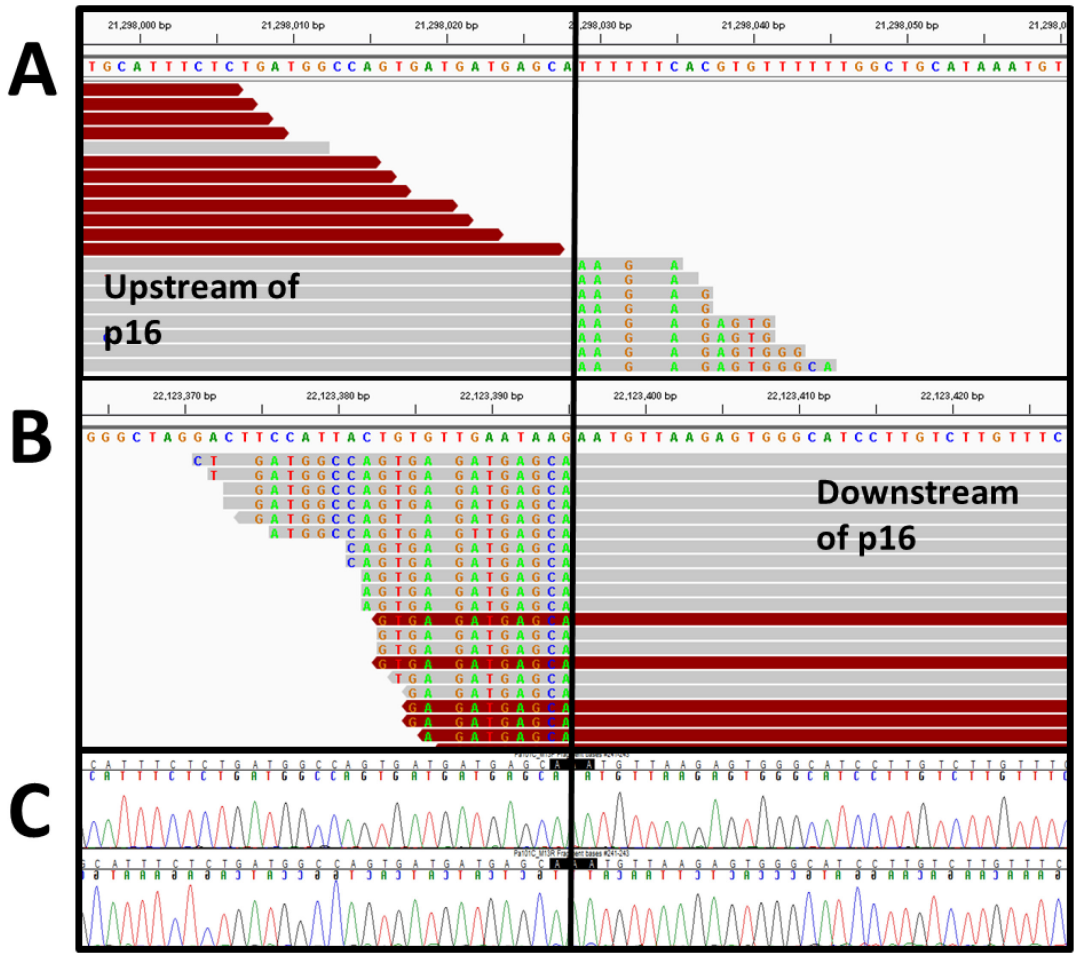


Figure 3.1. Signature of a HD in IGV. Shown is an interstitial deletion (0.8Mb) of *p16* in Pa101C (hg19: chr9:21,298,029; chr9:22,123,396), visualized with IGV and the reference sequence shown above each side of the breakpoint. The reads upstream (A) and downstream (B) of the *p16* HD are aligned, with the solid black line indicating the HD breakpoint. Colored bases indicate mismatched bases from reference, since they align to the other side of the breakpoint. Red-colored reads have an aberrant insert size with their pair, indicative of a HD. For this HD, there is no microhomology at the breakpoint and no non-templated bases inserted at the breakpoint. (C) Bidirectional Sanger sequencing confirms the novel junction.

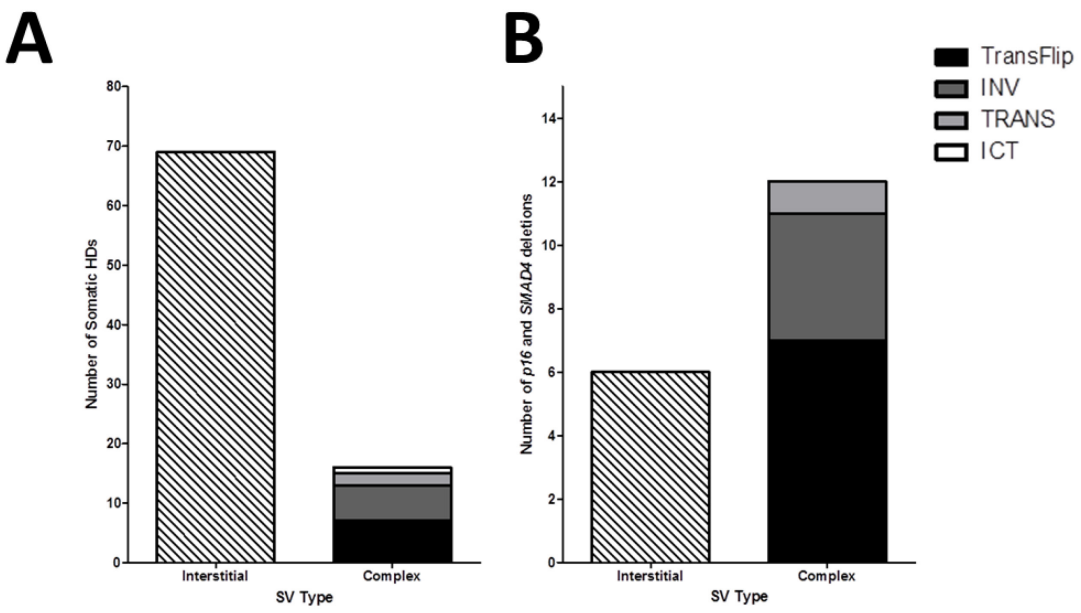


Figure 3.2. Most somatic HDs are simple interstitial deletions, but TransFlip mutations more prevalent in *p16* and *SMAD4* deletions. (A) Somatic HDs were categorized as interstitial, where breakpoints were directly end-joined, or complex structural variants (SV). The majority (69/85) of somatic HDs were interstitial. The 16 complex SVs were further characterized as TransFlip mutation (translocation at one breakpoint and inversion at the other breakpoint, 7/16), INV (inversion at both breakpoints, 6/16), TRANS (inter-chromosomal translocation at both breakpoints, 2/16), or ICT (intra-chromosomal translocation at both breakpoints, 1/16). (B) When looking at just *p16* and *SMAD4* deletions, the majority (12/18) are complex rearrangements, and most frequently TransFlip mutations (7/18). There are no ICT SVs underlying *p16* and *SMAD4* deletions.

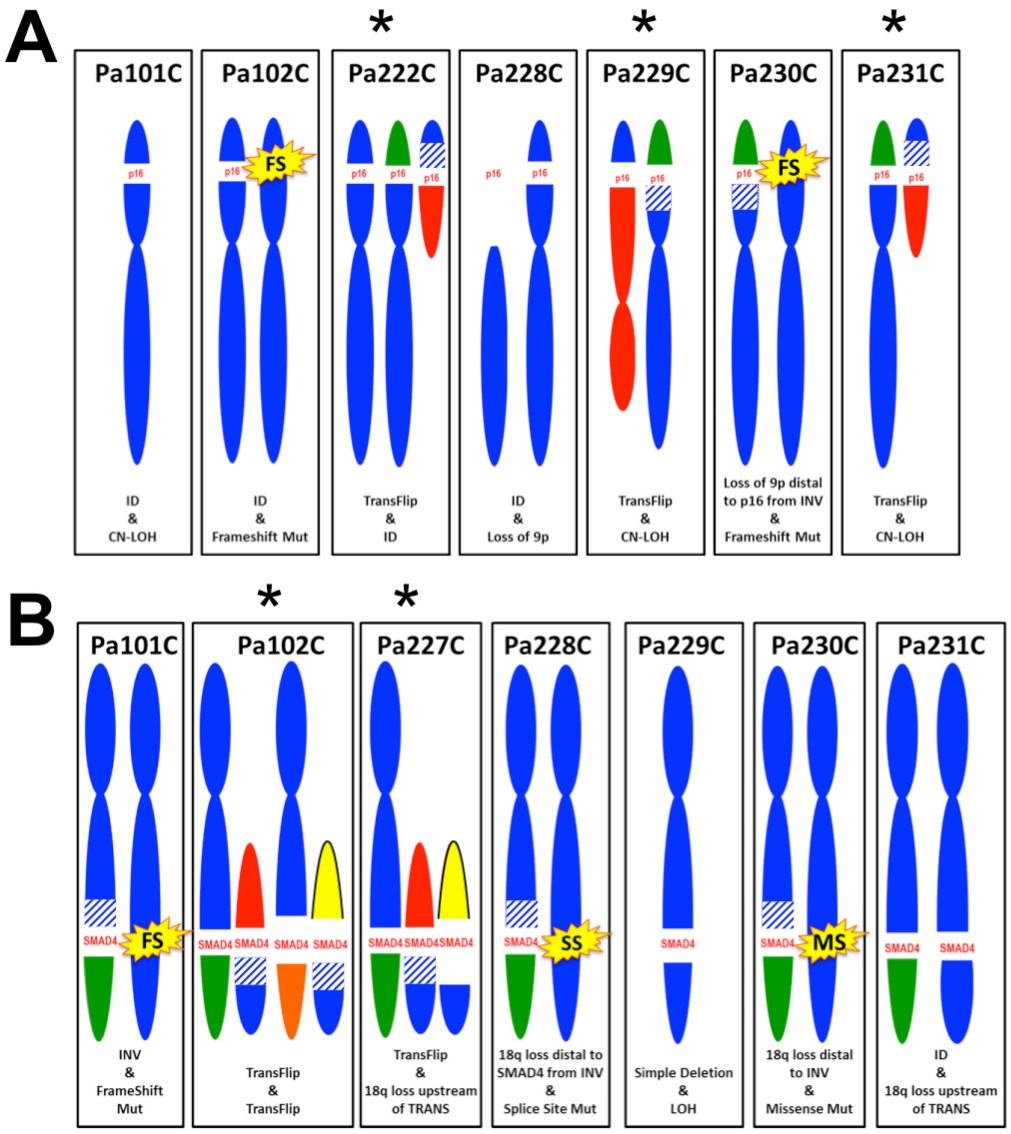


Figure 3.3. Derivative chromosomes for *p16* and *SMAD4* deletions. Shown are the derivative chromosomes 9 (A) and 18 (B). Deletions are indicated by gap of blue chromosome material, on the short arm of chromosome 9 for *p16* and the long arm of chromosome 18 for *SMAD4*. Translocations (TRANS) are indicated by non-blue-colored chromosome partners (all are non-recurrent), inversions (INV) are indicated by blue striping, point mutations (FS: frameshift, SS: splice site, MS: missense) are indicated by yellow starbursts. TransFlip mutations (*) underlie 3 *p16* deletions (Pa222C, Pa229C, and Pa231C), and 2 *SMAD4* deletions (Pa102C and Pa227C). Simple interstitial deletions (ID) and loss of heterozygosity (LOH, including copy-neutral events, or CN-LOH) also underlie *p16* and *SMAD4* deletions in our cell lines.

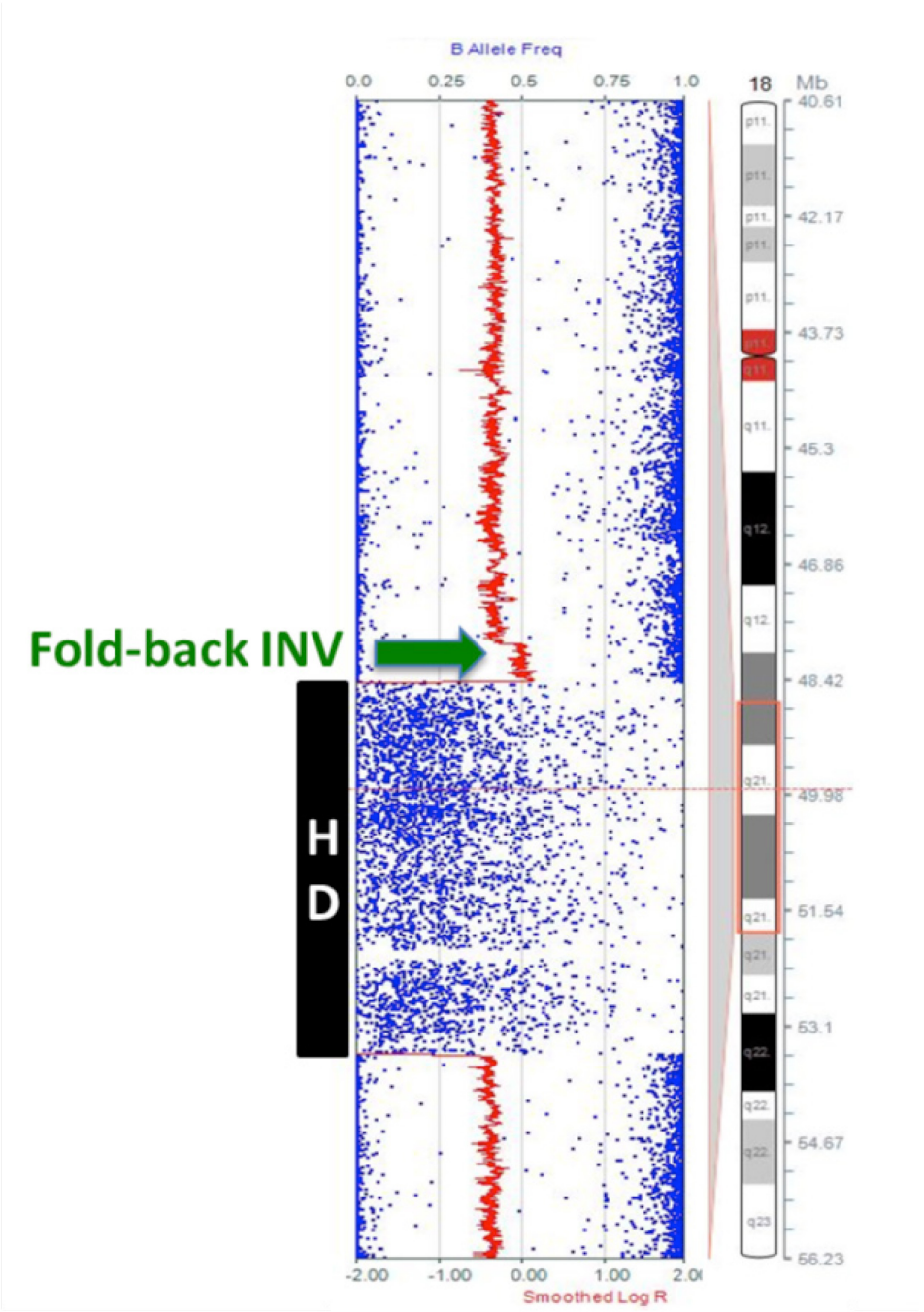


Figure 3.4. Fold-back inversion seen in one TransFlip mutation. High density SNP microarray of 18q21 in Pa222C reveals a copy number increase (green arrow) associated with the inversion side of the TransFlip mutation that results in the homozygous deletion of SMAD4 (black box). This is indicative of a fold-back inversion, and was not seen in any of the other inversions.

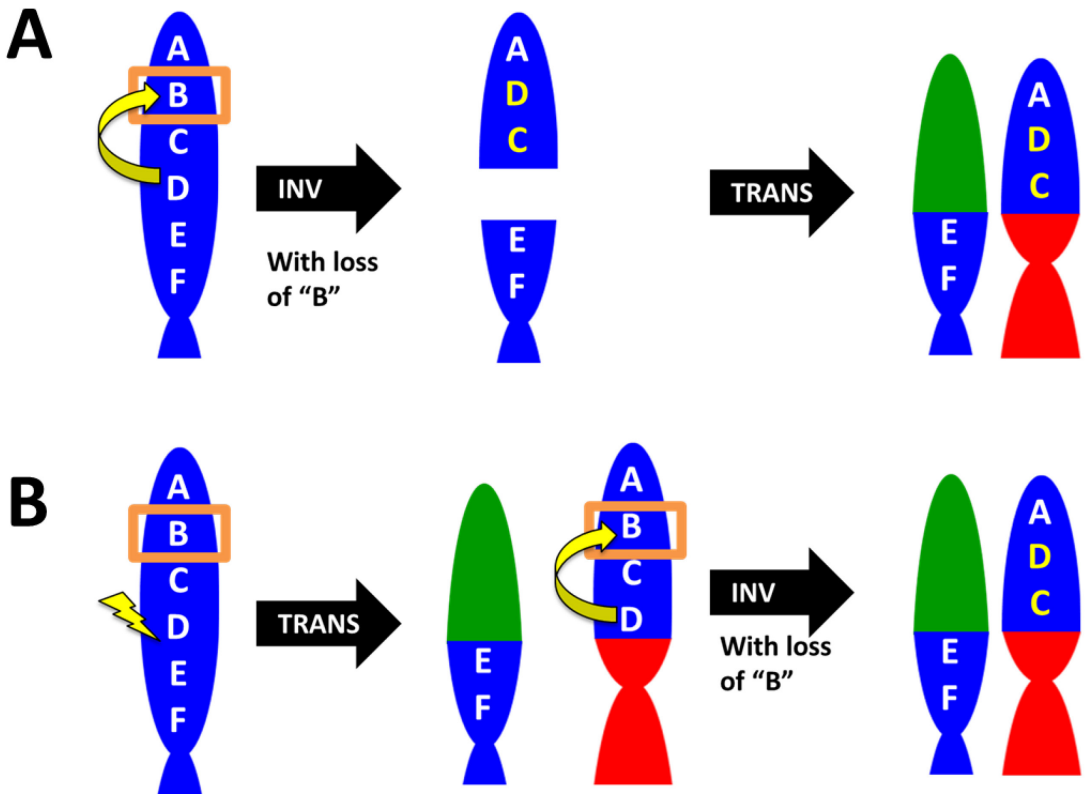


Figure 3.5. Proposed sequential mechanisms for TransFlip mutations. It is possible that the translocation and inversion occur simultaneously to resolve the deletion of B (boxed). Here we propose two stepwise mechanisms, with the TSG indicated by "B" on an example chromosome (blue). (A) An inversion of "BCD" (yellow) results in a DNA break and deletion of B, which is then resolved by a non-reciprocal translocation (red, green). (B) A DNA break between "D" and "E" (lightning bolt) is repaired by a non-reciprocal translocation (red, green) and followed by an inversion of "BCD" (yellow) and a deletion of "B".

Table 3.1: Summary of *p16* and *SMAD4* deletions

Gene	Sample	HD Size (Mb)	CNV	SV(s)
p16	Pa101C	0.8	2X†-HD-2X†	ID
	Pa102C	.	3X-1X-3X	ID
	Pa222C	0.1	1X-HD-1X	ID+TransFlip
	Pa227C*	.	.	.
	Pa228C	0.5	1X-HD-1X	ID[#]
	Pa229C	4	2X†-HD-2X†	TransFlip[#]
	Pa230C	.	1X-2X	INV[#]
	Pa231C	0.7	1X-HD-2X†	TransFlip[#]
SMAD4	Pa101C	.	3X-2X†	INV
	Pa102C	6	1X-HD-1X	TransFlip+TransFlip
	Pa222C*	.	.	.
	Pa227C	5	2X†-HD-1X	TransFlip[#]+TransFlip[#]
	Pa228C	.	3X-1X to qtel	INV[#]
	Pa229C	1.4	1X-HD-1X	ID[#]
	Pa230C	.	1X to qtel	INV[#]
	Pa231C	0.6	1X-HD-1X	ID+TRANS

2X†=Copy-Neutral LOH, **p16* is not deleted in Pa227C and *SMAD4* is not deleted in Pa222C, Dots (.) indicate that the structural or copy number alteration was not detected, [#]Confirmed in resected primary tumor samples.

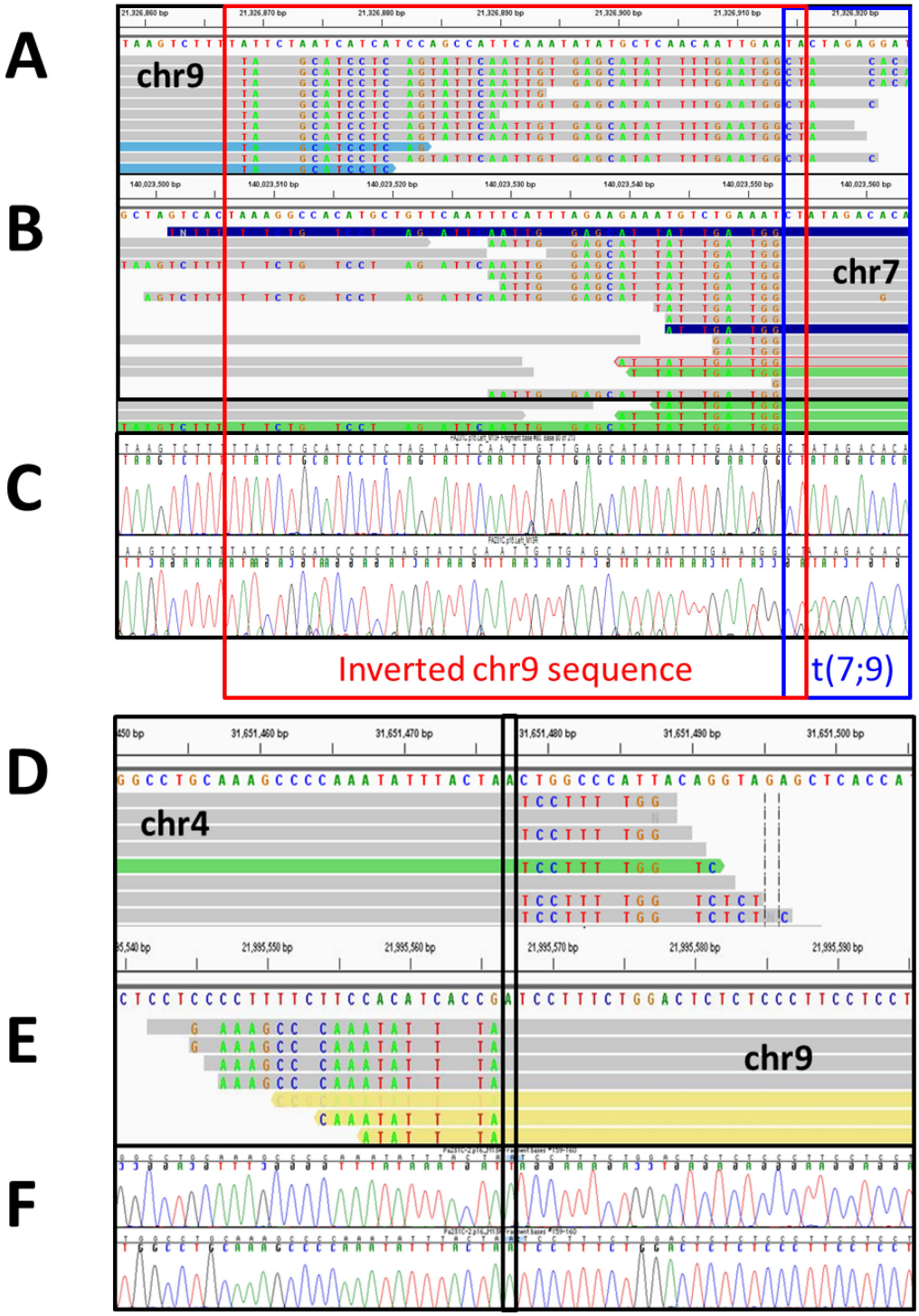
Table 3.2: Structural variants that underlie *p16* and *SMAD4* deletions

Tumor Suppressor Gene	Interstitial Deletion	TransFlip (TRANS+INV)	TRANS alone	INV alone
CDKN2A/p16	4	3	0	0
SMAD4/DPC4	2	4	1	3

Table 3.3: Characteristics of TransFlip mutations

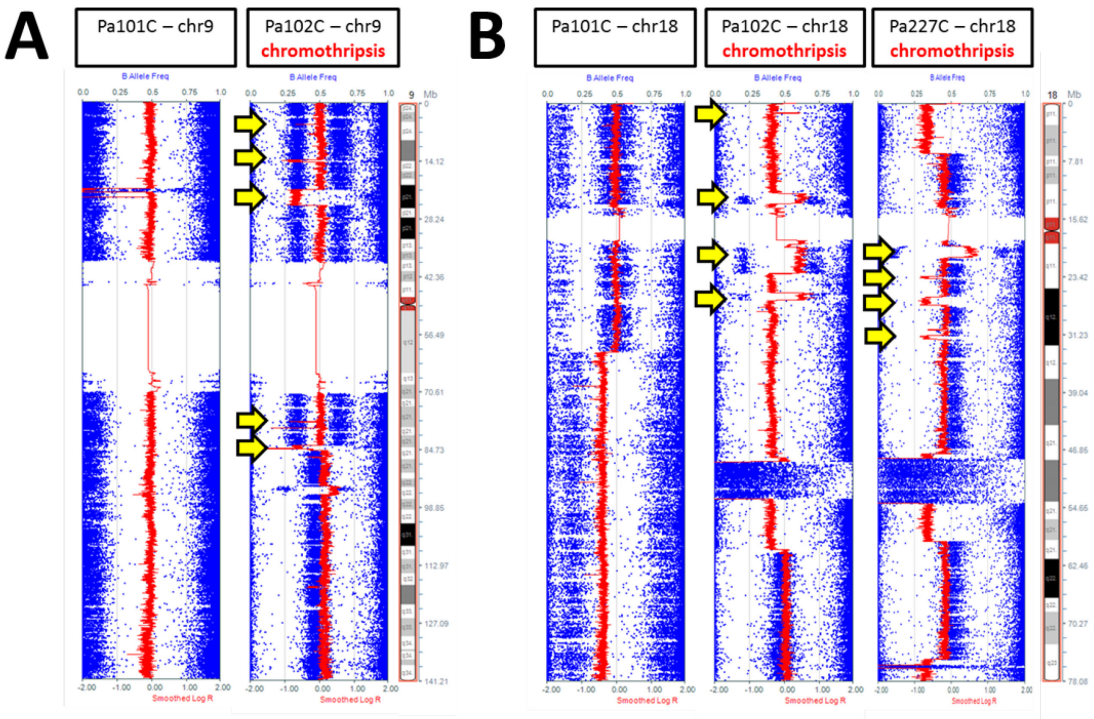
Characteristics of TransFlip Mutations
Non-recurrent breakpoints
Not present in the germline of these PDAC patients
Enriched at TSGs
Inversion size ranges from 47bp to 3.4kb
Fold-back inversions are uncommon
Inversions can be telomeric or centromeric to the affected TSG
Translocations are inter-chromosomal and non-reciprocal
Non-recurrent chromosome partners in translocations
Can occur with LOH, second TransFlip mutation, or an interstitial deletion to produce a TSG HD

Tumor suppressor genes (TSGs), loss of heterozygosity (LOH), homozygous deletion (HD)



Supplemental Figure 3.1. Complexity of TransFlip in IGV. The TransFlip mutation of *p16* in Pa231C involves a telomeric inversion (A-C) and centromeric translocation (D-F). Since the inversion is telomeric, it involves an additional translocation. (A-C) Here is shown the novel junctions created by the 49bp

inversion (hg19, chr9:21,326,868-chr9:21,326,916, red box) and translocation to chromosome 7 (hg19, chr7:140,023,553, blue box). The reads upstream, on chromosome 9 (A), and downstream, on chromosome 7 (B), of the translocation are aligned. The light blue-colored reads (A) indicate that their pairs are on chromosome 7, while the green-colored reads (B) indicate that their pairs are on chromosome 9. There are 2 bases of microhomology at the t(7;9) novel junction (overlap of red and blue boxes) and no non-templated bases inserted at the junction. (C) Bidirectional Sanger sequencing confirms the novel junctions. (D-E) The novel junction created by the translocation (hg19, chr4:31,651,477; chr9:21,995,567), with the reads upstream, on chromosome 4 (D), and downstream, on chromosome 9 (E), of the translocation aligned. The green-colored reads (D) indicate that their pairs are on chromosome 9, while the yellow-colored reads (E) indicate that their pairs are on chromosome 4. There is one base of microhomology at the breakpoint and no non-templated bases inserted at the junction. (F) Bidirectional Sanger sequencing confirms the novel junction.



Supplemental Figure 3.2. Chromothripsis in Pa102C and Pa227C. High density SNP microarray of chromothripsis (yellow arrows), where copy number switches between two states along the entire chromosome or chromosome arm. (A) Chromosome 9 exhibited chromothripsis in Pa102C. (B) Chromosome 18 exhibited chromothripsis in Pa102C and Pa227C. Pa101C chromosomes 9 and 18 do not exhibit chromothripsis and are shown for comparison.

Supplemental Table 3.1: Characteristics of known SV mechanisms

Mechanism	Characteristic(s)
NHEJ (non-homologous end-joining)	Insertions at junction Possibly short sequence homology
MMEJ (microhomology-mediated end joining)	Microhomology (overlying or flanking) Deletions
NAHR (non-allelic homologous recombination)	Recurrent SVs Repetitive sequence
MEI (mobile element insertion)	Intact transposable repetitive sequence (Alu, L1)
L1 3' transduction	Can mimic a translocation event polyA sequence flanking breakpoint
FoSTeS;MMBIR (fork stalling and template switching; microhomology-mediated break-induced repair)	Multiple complex rearrangements
Chromoplexy	Chain of rearrangements involving multiple chromosomes in areas of active transcription and open chromatin
Chromothripsis	Switching between 2 copy number states on a chrom.
BFB cycles (breakage-fusion-bridge)	Inversions and translocations; specifically fold-back inversions (increased copy number)
TransFlip mutations	Translocation on one side of a deletion and inversion on the other side of the deletion

Supplemental Table 3.2: Summary somatic HDs, by sample and by SV type

SAMPLE	INTERSTITIAL	COMPLEX				
		ALL Complex	Trans Flip	INV+ INV	TRANS+ TRANS	ICT+ ICT
Pa101C	3	0	0	0	0	0
Pa102C	11	5	2	2	1	0
Pa222C	5	0	0	0	0	0
Pa227C	6	1	1	0	0	0
Pa228C	10	2	1	1	0	0
Pa229C	14	4	2	1	0	1
Pa230C	5	1	0	1	0	0
Pa231C	15	3	1	1	1	0
Total	69	16	7	6	2	1
% of Total	81%	19%	8%	7%	2%	1%
median HD size (bp)*	9,379	153,470	668,700	57,180	584,589	1,134,599
min HD size (bp)	956	3,409	69,213	3,409	68,084	.
max HD size (bp)	1,417,741	6,045,187	6,045,187	389,465	1,101,094	.

*Complex are significantly larger, $p<0.05$; ICT=intra-chromosomal translocation

Supplemental Table 3.3: Details of all somatic HDs

Sample	SV Type	HD Size (bp)	HD Coordinates (hg19)	Genes
Pa101C	Interstitial HD	17,481	chr8:15,401,786-15,419,267	TUSC3
Pa101C	Interstitial HD	18,510	chr8:24,972,434-24,990,944	.
Pa101C	Interstitial HD	825,368	chr9:21,298,028-22,123,396	**CDKN2A/p16***, IFNA5, KLHL9, IFNA6, IFNA13, IFNA2, IFNA8, IFNA1, MIR31HG, IFNE, MIR31, MTAP, CDKN2A-AS1, CDKN2B, CDKN2B-AS1
Pa102C	Interstitial HD	1,126	chr6:16,884,968-16,886,094	.
Pa102C	Interstitial HD	1,633	chr3:65,117,912-65,119,545	.
Pa102C	Interstitial HD	5,710	chr3:98,943,806-98,949,516	.
Pa102C	Interstitial HD	6,959	chr8:40,182,856-40,189,815	.
Pa102C	Interstitial HD	8,157	chr6:19,041,219-19,049,376	.
Pa102C	Interstitial HD	9,746	chr8:15,777,416-15,787,162	.
Pa102C	Interstitial HD	11,828	chr3:53,027,071-53,038,899	SFMBT1, RFT1, PRKCD, TKT, DCP1A, Y_RNA, SNORA26
Pa102C	Interstitial HD	18,510	chr8:24,972,434-24,990,944	.
Pa102C	Interstitial HD	48,304	chr5:97,047,855-97,096,159	.
Pa102C	Interstitial HD	54,259	chr5:180,376,278-180,430,537	BTNL3, BTNL8
Pa102C	Interstitial HD	155,155	chr8:39,232,074-39,387,229	ADAM5P, ADAM3A
Pa102C	INV+INV	92,568	chr11:55,365,428-55,457,996	OR4C11, OR4P4, OR4S2
Pa102C	INV+INV	389,465	chr18:75,924,436-76,313,901	.
Pa102C	TRANS+ TRANS	68,084	chr9:84,607,953-84,676,037	FAM75D1
Pa102C	TransFlip	69,213	chr1:248,741,764-248,810,977	OR2T10, OR2T11
Pa102C	TransFlip	6,045,187	chr18:47,999,471-54,044,658	**SMAD4**[#], MAPK4, MRO, ME2, ELAC1, MEX3C, LOC100287225, DCC, LOC102724651, LOC101928167, MBD2, SNORA37, POLI, STARD6, C18orf54, DYNAP, RAB27B, CCDC68, LOC101927229, TCF4, MIR4529, LOC101927273, LINC01539
Pa222C	Interstitial HD	1,005	chr3:61,127,189-61,128,194	FHIT
Pa222C	Interstitial HD	9,752	chr4:10,392,430-10,402,182	.

Pa222C	Interstitial HD	10,837	chr12:59,935,796-59,946,633	.
Pa222C	Interstitial HD	34,123	chr8:83,740,865-83,774,988	.
Pa222C	Interstitial HD	97,053	chr9:21,977,376-22,074,429	**CDKN2A/p16***, CDKN2B, CDKN2B-AS1
Pa227C	Interstitial HD	3,794	chr6:8,337,733-8,341,527	.
Pa227C	Interstitial HD	6,332	chr6:141,751,266-141,757,598	.
Pa227C	Interstitial HD	8,488	chr8:104,271,982-104,280,470	LOC100499183
Pa227C	Interstitial HD	9,441	chr14:35,605,628-35,615,069	KIAA0391
Pa227C	Interstitial HD	12,776	chr17:54,160,182-54,172,958	.
Pa227C	Interstitial HD	18,510	chr8:24,972,434-24,990,944	.
Pa227C	TransFlip	5,001,960	chr18:48,463,089-53,465,049	**SMAD4**[#], ME2, ELAC1, MEX3C, LOC100287225, DCC, LOC102724651, LOC101928167, MBD2, SNORA37, POLI, STARD6, C18orf54, DYNAP, RAB27B, CCDC68, LOC101927229, TCF4, MIR452
Pa228C	Interstitial HD	956	chr6:74,832,504-74,833,460	AF086303
Pa228C	Interstitial HD	2,442	chr6:24,325,366-24,327,808	DCDC2
Pa228C	Interstitial HD	3,667	chr18:73,862,188-73,865,855	.
Pa228C	Interstitial HD	3,746	chr2:195,363,696-195,367,442	.
Pa228C	Interstitial HD	4,606	chr18:51,205,969-51,210,575	.
Pa228C	Interstitial HD	6,113	chr17:56,207,497-56,213,610	.
Pa228C	Interstitial HD	6,131	chr17:70,815,230-70,821,361	SLC39A11
Pa228C	Interstitial HD	6,465	chr7:110,181,968-110,188,433	.
Pa228C	Interstitial HD	7,652	chr6:48,930,882-48,938,534	.
Pa228C	Interstitial HD	485,050	chr9:21,950,541-22,435,591	**CDKN2A/p16***, CDKN2B, CDKN2B-AS1
Pa228C	INV+INV	21,791	chr9:22,720,074-22,741,865	FLJ35282
Pa228C	TransFlip	205,910	chr9:6,650,981-6,856,891	KDM4C, DQ580140
Pa229C	ICT+ICT	1,134,599	chr17:63,527,925-64,662,524	AXIN2, CEP112, APOH, PRKCA
Pa229C	Interstitial HD	1,007	chr8:17,580,760-17,581,767	MTUS1
Pa229C	Interstitial HD	3,634	chr4:152,990,313-152,993,947	.

Pa229C	Interstitial HD	4,625	chr18:63,907,136-63,911,761	.
Pa229C	Interstitial HD	4,759	chr5:111,939,582-111,944,341	.
Pa229C	Interstitial HD	4,904	chr4:112,237,509-112,242,413	.
Pa229C	Interstitial HD	6,107	chr1:62,113,414-62,119,521	.
Pa229C	Interstitial HD	6,957	chr18:50,242,911-50,249,868	DCC
Pa229C	Interstitial HD	6,959	chr8:40,182,856-40,189,815	.
Pa229C	Interstitial HD	8,527	chr18:63,723,838-63,732,365	.
Pa229C	Interstitial HD	9,379	chr22:30,044,155-30,053,534	NF2
Pa229C	Interstitial HD	15,381	chr18:50,014,115-50,029,496	DCC
Pa229C	Interstitial HD	19,143	chr4:189,248,699-189,267,842	.
Pa229C	Interstitial HD	24,080	chr13:69,244,702-69,268,782	.
Pa229C	Interstitial HD	1,417,741	chr18:48,434,141-49,851,882	**SMAD4**[#], ME2, ELAC1, MEX3C, LOC100287225
Pa229C	INV+INV	10,274	chr4:116,166,907-116,177,181	.
Pa229C	TransFlip	87,849	chr22:36,250,342-36,338,191	RBFOX2
Pa229C	TransFlip	3,975,324	chr9:20,345,146-24,320,470	**CDKN2A/p16***, MLLT3, MIR4473, MIR4474, FOCAD, FOCAD-AS1, MIR491, PTPLAD2, IFNB1, IFNW1, IFNA21, IFNA4, IFNA7, IFNA10, IFNA16, IFNA17, IFNA14, IFNA22P, IFNA5, KLHL9, IFNA6, IFNA13, IFNA2, IFNA8, IFNA1, MIR31HG, IFNE, MIR31, MTAP, CDKN2A-AS1, CDKN2B, CDKN2B-AS1, DMRTA1, LINC01239, LOC101929563, ELAVL2
Pa230C	Interstitial HD	4,478	chr3:6,650,160-6,654,638	AF279782
Pa230C	Interstitial HD	4,618	chr3:11,410,124-11,414,742	ATG7
Pa230C	Interstitial HD	4,886	chr6:132,707,536-132,712,422	MOXD1
Pa230C	Interstitial HD	10,138	chr9:6,700,475-6,710,613	.
Pa230C	Interstitial HD	31,893	chr13:24,631,706-24,663,599	SPATA13
Pa230C	INV+INV	3,409	chr11:97,517,721-97,521,130	.
Pa231C	Interstitial HD	2,340	chr9:79,555,762-79,558,102	.
Pa231C	Interstitial HD	5,146	chr6:77,097,496-77,102,642	.

Pa231C	Interstitial HD	5,318	chr7:159,117,444-159,122,762	.
Pa231C	Interstitial HD	7,898	chr4:41,969,874-41,977,772	.
Pa231C	Interstitial HD	9,858	chr6:81,283,720-81,293,578	.
Pa231C	Interstitial HD	10,600	chr3:12,904,010-12,914,610	DQ587889, DQ587809
Pa231C	Interstitial HD	17,028	chr9:24,502,071-24,519,099	.
Pa231C	Interstitial HD	20,055	chr4:64,694,328-64,714,383	.
Pa231C	Interstitial HD	21,792	chr6:77,437,256-77,459,048	.
Pa231C	Interstitial HD	64,781	chr8:123,562,465-123,627,246	.
Pa231C	Interstitial HD	107,148	chr8:138,259,213-138,366,361	.
Pa231C	Interstitial HD	134,333	chr6:91,985,925-92,120,258	.
Pa231C	Interstitial HD	498,258	chr4:178,588,820-179,087,078	AK094945, LOC285501
Pa231C	Interstitial HD	567,944	chr3:69,892,576-70,460,520	MITF, BC015590
Pa231C	Interstitial HD	621,563	chr18:48,570,319-49,191,882	**SMAD4**[#], MEX3C, LOC100287225
Pa231C	INV+INV	101,030	chr12:28,931,318-29,032,348	.
Pa231C	TRANS+TRANS	1,101,094	chr17:11,169,977-12,271,071	SHISA6, DNAH9, ZNF18, MAP2K4, MIR744, U11
Pa231C	TransFlip	668,700	chr9:21,326,867-21,995,567	**CDKN2A/p16***, KLHL9, IFNA6, IFNA13, IFNA2, IFNA8, IFNA1, MIR31HG, IFNE, MIR31, MTAP, CDKN2A-AS1, CDKN2B, CDKN2B-AS1

Homozygous deletion (HD), structural variant (SV), translocation (TRANS), inversion (INV), **p16* deletion breakpoints, [#]*SMAD4* deletion breakpoints

Supplemental Table 3.4: *p16* and *SMAD4* deletions details

Gene	Sample	HD Size (Mb)	CNV Pattern	Underlying SV for 1st Del (producing 1X/LOH)	Underlying SV for 2nd Del (producing HD)
p16	Pa101C	0.8	2X†-HD-2X†	CN-LOH (chr9 is 2X†)	Interstitial Deletion
	Pa102C	.	3X-1X-3X	Interstitial Deletion (chr9 is 3X)	.
	Pa222C	0.1	1X-HD-1X	**159bp INV with t(3;9)(p12.3;p21.3)** / **t(9;10)(p21.2;q26.3)**	Interstitial Deletion
	Pa228C	0.5	1X-HD-1X	LOH (9p is 1X)	Interstitial Deletion
	Pa229C	4	2X†-HD-2X†	CN-LOH (chr9 is 2X†)	**t(9;22)(p21.3;q12.3)** / **3,373bp INV with t(9;22) (p21.3;q12.3)**
	Pa230C	.	1X-2X	INV	.
	Pa231C	0.7	1X-HD-2X†	LOH (1X from HD to 9ptel)	**INV with t(7;9)(q34;p21.3)** / **t(4;9)(p15.1;p21.3)x2**
SMAD4	Pa101C	.	3X-2X†	28.5Mb INV	.
	Pa102C	6	1X-HD-1X	**t(14;18)(q21.2;q21.1)** / **INV to repetitive region**	**t(2;18)(q36.1;q21.1)** / **35Mb INV to centromere**
	Pa227C	5	2X†-HD-1X	**INV to repetitive region** / **t(17;18)(q21.31;q21.33)**	**200bp INV** / **t(6;18)(q22.31;q21.2)**
	Pa228C	.	3X-1X to qtel	264bp INV	.
	Pa229C	1.4	1X-HD-1X	1X to centromere, where becomes 3X to 18ptel / 1X to 18qtel	Interstitial Deletion
	Pa230C	.	1X to qtel	53bp INV	.
	Pa231C	0.6	1X-HD-1X	**t(5;18)(p14.3;q11.2)** / 1X to 18qtel	Interstitial Deletion

Homozygous deletion (HD), copy number variant (CNV), structural variant (SV), loss of heterozygosity (LOH; 1X), deletion (Del), copy-neutral LOH (CN-LOH; 2X†), inversion (INV),

Supplemental Table 3.5: Patient and cell line characteristics

Cell Line	Patient Age	Patient Sex	Passage Number	Karyotype
Pa101C	40s	M	P10	na
Pa102C	60s	F	P12	51~59<3n>,X,-X,-X,+1,add(1)(q32),der(1;7)(p10;p10), der(1;19)(q10;p10)add(1)(q32),add(2)(p11.2),add(3)(p11),-5,-6, add(6)(p22),add(7)(q21), add(8)(q24)x2,i(8)(q10),-9,-9,add(9)(q21),add(10)(p11.2), add(11)(p12),der(13;16)(q10;p10),-14,-14,-15,psu idic(15)(p13),+der(16)t(15;16)(q15;q22), add(17)(p11.2),+i(17)(q10),-18,-18,add(18)(q23),add(19)(p13.3), idic(20)(p11.2),-21,-22,-22,+mar1,+mar2,+mar3[cp22]
Pa222C	70s	M	P14	66~73<3n>,XY,add(X)(p11.1),+der(1)t(1;4)(p13:q12),del(2)(q21q23),?inv(2)(q22q23),add(3)(p12),add(4)(p13),der(6)t(5;6)(q13;q16), +7,del(8)(q13q22),-9,add(9)(p12),del(10)(q22),-12,+13, der(14)t(3;14)(q11.2;p11.2),+16,+17,der(17;22)(q10;q10)x2,-18, -19,+20,-21,+4~6mar[22]
Pa227C	60s	F	P12	38,X,-X,der(1)del(1)(p34)t(1;14)(q43;q11.2), der(5;10)(q10;q10),add(6)(p22),add(6)(q21),-8,add(9)(q13),-13, add(13)(p11.2), i(14)(q10),-17,-18,-18,-19,-20,-21,-22,+4mar[20]
Pa228C	60s	F	P9	70,XX,-X,+der(1;3)(p10;q10),del(2)(q33),del(3)(p13),-4,+5,+i(7)(p10),add(9)(p11),add(9)(p21)x2,-13,add(14)(p13), -18,+add(19)(p13.2),+20,add(20)(p11.2)x2,-21,-22,+2mar[13]/68~71,sl,+del(2)(q33),+4,-5,+6,-i(7)(p10),-8, +add(9)(p11),-10,+11.+13,+add(14)(p13),-15,-19,-add(19)(p13.2), +der(19)add(19)(p13.1)hsr(19)(p13.1),+21,+mar[cp7]
Pa229C	40s	M	P15	54~71<3n>,XX,-Y,+add(1)(p13),+2,-4,-5,+6,add(7)(q11.1),+add(7)(q22)x2,der(8;14)(q10;q10),+der(8;14), -9,add(9)(p12)x2,-10,add(10)(p15)x2,add(12)(p13), add(12)(q24.3),-13,-14,-15,-16,add(16)(q24),-17,-17,-17,-18, -18,-18,i(19)(q10),+20,?i(21)(q10),-22,-22,add(22)(q13),+r,+9~12mar[cp18]
Pa230C	60s	F	P11	71~86<4n>,XX,-X,-X,-1,-1,add(3)(p13)x2,add(3)(p22)x2,-5,-5,add(5)(p13),i(5)(p10),-6,-6,add(7)(p22)x2,-8,i(8)(q10)x2,-9,-9,-11,del(11)(p11.2)x2,-12,add(12)(p11.2)x2,-13,-13,-14,-15,add(16)(p13.3)x2,add(16)(q12),-17,-17,-18,-18,add(19)(q13.3)x2,-20,-20,-21,-21, add(21)(p11.2)x2,der(?)t(?;9)(?;q13)x2,mar1x2,mar2,mar3[cp19]
Pa231C	60s	M	P12	62~68<3n>,XY,-X or -Y,del(1)(p21p31),-2,i(3)(q10),-4,+5,del(5)(q12q14)x2,-6,+7,+8,der(8;22)(q10;q10)x2, -9,**der(9)t(4;9)(p15;p21)x2**[#],+11,-13,-14,-15,del(15)(q22),add(17)(p11.1),add(17)(p11.2)x2,-18,-18, ?i(19)(q10),+20,+20,-21,-21, i(21)(q10),+mar1x3,+mar2,+mar3[cp20]

Supplemental Table 3.6: Breakpoint characteristics of *p16* and *SMAD4* deletions

Gene	Sample	SV	Potential Mechanism						
			NHEJ	**MHEJ**	**MEI**	**L1 3' Trans.**	**Chromoplexy**	**Chromo-thripsis**	**Fragile Site**
			Ins Bases	MH at jxn	RepeatMasker at Jxn	Ins polyA at jxn	Genes within 10kb	Switching Copy#	At TRANS chr
p16	Pa101C	ID	0	0	**L1PA4 (L); L1MA5A (R)**	N	N	N	.
	Pa102C	ID	0	**1**	**MERV50-int (L); L1M3c (R)**	N	N	**Y**	.
	Pa222C	INV with TRANS	0/0	**4/1**	**L1ME3A (INV); L2a (3)**	N/N	**ROBO1 (3)**	N	N
		ID	0	**3**	**AluSx (L)**	N	**p16 (L), CDKN2B-AS1 (R)**		.
		TRANS	0	**1**	**L1MEf (10); MSTB-int (9)**	N	N		N
	Pa228C	ID	0	**2**	**AluSp (L); HERV9-int (R)**	N	**MTAP (L)**	N	.
	Pa229C	TRANS	0	**1**	N	N	**MLLT3 (L)**	N	**Y**
		INV	0	**2**	**MSTB-int (L)**	N	N		.
	Pa230C	INV-1	0	0	N	N	N	N	.
		INV-2	0	**3**	**LTR1D (R)**	N			
	Pa231C	INV with TRANS	0/0	**1/7**	**L1MB8 (7)**	N/N	**KLHL9 (9)**	N	N
		TRANS	0	**1**	**MER58B (4)**	N	**MTAP (9)**		**Y**
SMAD4	Pa101C	INV	0	0	**L2c (L)**	N	**RPRD1A (L)**	N	.
	Pa102C	TRANS	0	**3**	N	N	**SNHG22 (18); LINC00871 (14)**	**Y**	N
		INV	?	?	?	?	?		.
		TRANS	0	0	**MSTD (2)**	N	N		N
		INV	**1**	**1**	**L1MC4a (R)**	N	N		.
	Pa227C	fold-back INV	0	**1**	**AluJr4 (L); L1MEg (R)**	N	**ME2 (L&R)**	**Y**	.
		TRANS	0	**3**	**L1MEc (18)**	N	**NKAIN2 (6)**		N
		INV	?	?	?	?	?		.
		TRANS	0	0	**L1M3 (18)**	N	**ADAM11 (17)**		N
	Pa228C	INV	0	**1**	**L3 (L&R)**	N	N	N	.
	Pa229C	ID	0	0	**L1M5 (L); AluSp (R)**	N	**ME2 (L)**	N	.

Pa230C	INV	0	**1**	N	N	N	N	.
Pa231C	ID	0	0	**L2b (R)**	N	**SMAD4 (L)**	N	.
	TRANS	0	**3**	**TcMar-Tigger (18)**	N	**TAF4B (18); CDH12 (5)**		**Y**

Interstitial deletion (ID), inversion (INV), translocation (TRANS), left breakpoint (L), right breakpoint (R), chromosome number (#), microhomology (MH), non-homologous end-joining (NHEJ), microhomology end-joining (MHEJ), mobile element insertion (MEI), LINE-1 3' transduction (L1 3' Trans.)

Chapter 4

Challenges in discovering familial pancreatic cancer predisposition genes

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is a nearly uniformly lethal disease, where up to 10% of the cases occurring in familial aggregates. The genes underlying the majority (>80%) of these familial pancreatic cancer (FPC) cases are unknown. Previous whole exome sequencing (WES) strategies have successfully identified FPC predisposition genes *PALB2* and *ATM*, which allows for the identification of at-risk individuals for which early detection resources can be focused on. In this study, we have employed an integrative strategy of high density SNP microarray, WES, whole genome sequencing (WGS), and RNA-Seq to identify new FPC predisposition genes

Introduction

Over 48,000 cases of pancreatic ductal adenocarcinoma (PDAC) are expected to be diagnosed in the U.S. in 2015, and nearly 5,000 will occur in familial aggregate (familial pancreatic cancer, FPC) (*92*). FPC is defined as a pancreatic cancer in a patient with at least one first-degree relative with pancreatic cancer (i.e. 2 or more affected first-degree relatives). A challenge in early detection efforts for FPC is identifying at-risk individuals since the vast majority of FPC kindreds have an unidentified predisposition gene. Currently known FPC predisposition genes, identified by traditional linkage mapping and more recently, next generation sequencing, account for less than 20% of FPC cases (see Table 1.1). Recent efforts have highlighted the success of sequencing studies to identify new FPC predisposition genes, with *PALB2* and *ATM* discovered using whole exome sequencing (WES) of the tumor and matched normal for FPC patients (*17, 93*).

Most cancer predisposition genes act as tumor suppressors (TSGs) with biallelic inactivating alterations, following the two hit hypothesis first described by Nicholls in 1969 and popularized by Knudson (*26, 27*). In the two hit model for inherited cancer, the first hit (mutation) in the TSG is inherited and the second hit in the same TSG is acquired somatically by chance. Whereas oncogenes are most commonly mutated at recurrent same amino acid positions ("hotspots"), TSGs are mutated through protein truncating mutations that occur most often throughout the length of the protein (*94*).

Here, we build upon the successful strategy employed to identify the *PALB2* and *ATM* genes. We used an integrated approach including high density SNP microarrays, exomic sequencing (WES), whole genome sequencing (WGS), and RNA-sequencing (RNA-Seq) to identify new candidate FPC predisposition genes. The majority of our screening was performed under the assumption that our FPC gene would conform to the classic two-hit hypothesis.

Materials and Methods

Study Participants

This study was reviewed and approved by the Institutional Review Board at Johns Hopkins Medical Institutions, and informed consent was obtained from all study participants. Cancer cell lines were generated from the tumors of patients with FPC, obtained from resection surgery or rapid autopsy, and generated by direct culture or xenograft, as previously described (*68*). Cell lines were karyotyped using a previously described method (*69*). Matched normal DNA was obtained from either frozen tissue or Epstein-Barr virus (EBV)-transformed lymphoblasts.

Preparation of Genomic DNA and RNA

Genomic DNA was extracted from early passage cell lines and matched normal using QIAamp DNA mini kit (Qiagen, Valencia, CA), per manufacturer's instruction. All cell lines were verified by STR analysis using the ABI Profiler kit (Life Technologies, Carlsbad, CA) on the 3130xL Genetic Analyzer instrument (Life Technologies). RNA was extracted from cell lines using RNeasy mini kit (Qiagen), per manufacturer's instruction. A HPDE (human pancreatic ductal epithelium) cell line was used as a normal control for RNA-Seq.

High Density SNP Microarray

The Omni2.5 array (Illumina, San Diego, CA) was used to analyze cancer cell lines and matched normal samples at 2,379,855 (2.5M) SNP loci (one SNP every ~1,000bp). GenomeStudio (Illumina) was used to identify copy number variants (CNVs) using the following criteria: an average LogR Ratio (LRR) $\leq$ -2.0 for homozygous deletions; LRR of 0-0.53 and B Allele Frequency (BAF) of 0 or 1 for loss of heterozygosity (LOH); and an average LRR $\geq$ 1.4, with at least 1 SNP LRR $\geq$ 2.0, for amplifications. A minimum of 4 SNPswere required for the region to be called an alteration, with the boundaries being the first and last SNPs that meet criteria, and adjacent regions (within 100kb) were considered to be one alteration. KaryoStudio (Illumina) was used to visually confirm CNVs.

Whole exome and whole genome sequencing

For whole exome sequencing, genomic DNA libraries prepared with Agilent's SureSelect Paired-End Version 2.0 Human Exome Kit (Agilent, Santa Clara, CA) were paired-end sequenced (2x75bp) on a GAIIx Genome Analyzer (Illumina) to 200X coverage (*68*). For whole genome sequencing, genomic DNA was paired-end sequenced (2x100bp) on a HiSeq 2000 (Illumina) to a 60X coverage for cancers and 30X coverage for matched normal by Personal Genome Diagnostics (Baltimore, MD). For both WES and WGS, the resulting reads were aligned to human genome (hg19) with Eland v.2 algorithm in CASAVA 1.7 software (Illumina).

RNA-Seq

Using TruSeq Stranded Total RNA Sample Preparation (Illumina), cDNA libraries were prepared from ribosomalRNA-depleted Total RNA and paired-end sequenced (2x100bp) on a HiSeq 2000 to a level of 50M reads. RSEM was used to align the sequences to human genome (hg19) (*31*). Alterations were visually confirmed using IGV.

Variant Filtering

Since we were looking for TSGs, we specifically looked for private truncating variants (PTVs), which include copy number deletions, frameshifting insertions and deletions, stop gain mutations, stop loss mutations, missense mutations of the first amino acid, and splice site mutations in positions -1, -2, +1, or +2 that are not common in the general population (<1% minor allele frequency, MAF). We annotated mutations using ANNOVAR (60), to include information about MAF, predicted effect on protein (PolyPhen, SIFT), and conservation (phastConsElements 46-way).

Our strategy looked for genes with both an alteration in cancer cell line (somatic hit) and an alteration in the matched normal sample (germline hit). The hits could be large inactivating alterations (e.g. deletions) or small inactivating alterations (e.g. point mutations). Candidate mutations were first visually confirmed using Integrated Genomics Viewer (IGV) (Broad

Institute, Boston, MA) and then experimentally confirmed by PCR amplification and bidirectional Sanger sequencing.

Results

Integrated Approach

In this chapter, we carried out high density SNP microarray (probes every ~1,000 bases), WES, WGS, RNA-Seq, and conventional karyotyping (Figure 1). We looked for deletions by SNP microarray; for point mutations by WES and WGS, and for loss of expression by RNA-Seq. The methods also served to cross-confirm mutations or deletions discovered. Bidirectional Sanger sequencing was used to confirm candidate mutations and deletions.

Germline intragenic HDs in Both FPC and Normal Controls

We first investigated whether the FPC patients might carry HDs in predisposition genes. The number of germline intragenic HDs in FPC samples were not statistically increased from normal controls (Figure 2, Table 2) (median 5 vs 4, p=0.20 by unpaired two-tailed t-test), and the amount of genetic material deleted in our FPC samples was also not statistically increased from normal controls (median 68kb vs 90kb, p=0.64 by unpaired two-tailed t-test). The locus including *ADAM5P* and *ADAM3A* was recurrently deleted in the germline of FPC patients (PA228, PA229, and PA231). However, analysis of normal controls showed this specific HD is common in the general population, and thus likely benign. All germline deletions were simple interstitial deletion, with the exception of a recurrent extragenic deletion with an inversion at each breakpoint.

No Candidate Genes Identified by Germline PTV Analysis

No clearly deleterious mutations were identified in *ATM*, *STK11*, *PRSS1*, *PALB2*, *BRCA2*, or *SPINK1*. No loss of expression was seen in the genes either. We looked for genes with a germline defect in combination with a somatic defect ("2nd hit"). The resulting hits from these either did not confirm (e.g. CDC27, result of pseudogenes) or were found to be common (and thus likely benign) variants by exome variant server (ESP6500) and 1000 Genomes databases (MAF>1%). Since DNA damage repair genes account for the majority of known FPC

predisposition genes, we extended our search to missense mutations with MAF<1% and predicted to be deleterious by both PolyPhen and SIFT, but still had no genes with biallelic inactivation, where one defect was in the germline. Possible dominant negative candidate genes with suggestive 1st hits, but lack of an obvious 2nd hit, include *STK11IP* (Pa229C) and *RBBP8* (Pa230C), which may warrant follow-up functional studies.

Discussion

Here we have characterized a cohort of 8 FPC patients using an integration of chromosome-level, DNA-level, and RNA-level analyses. Despite our detailed analysis and follow-up testing of many putative genes, we have no bona fide candidate gene for FPC predisposition. Our research does highlight that normal germlines have homozygous deletions of genes, like *ADAM5P* and *ADAM3A*. It remains to be seen if there is any biological consequence for these deletions.

Current analysis efforts are focused on RNA-Seq level variation, including allele-specific expression (ASE), which may indicate epigenetic silencing of the second allele (2^{nd} hit), and differential gene expression between our FPC cohort and a sporadic PDAC cohort. The possible role of alternate or aberrant splicing is also being explored.

Future work could add proteomics and epigenetics analyses of this cohort of FPC lines. However that still does not guarantee the identification of a predisposition gene. Instead, we investigators need to explore novel mechanisms of cancer predisposition, beyond biallelic inactivation of a gene by intragenic mutations. Future analysis should explore (1) monoallelic inactivation, such as haploinsufficiency and dominant negative mechanisms, and (2) intergenic or non-coding mutations that may produce dysregulated gene expression.

It is possible that these FPC kindreds of our cohort do carry an inherited defect that predisposes to PDAC, but the specific case in each kindred that we have studied does not carry the inherited defect and are just coincidental sporadic PDAC cases in FPC kindreds, the definition of "phenocopies." We find this exceedingly improbable.

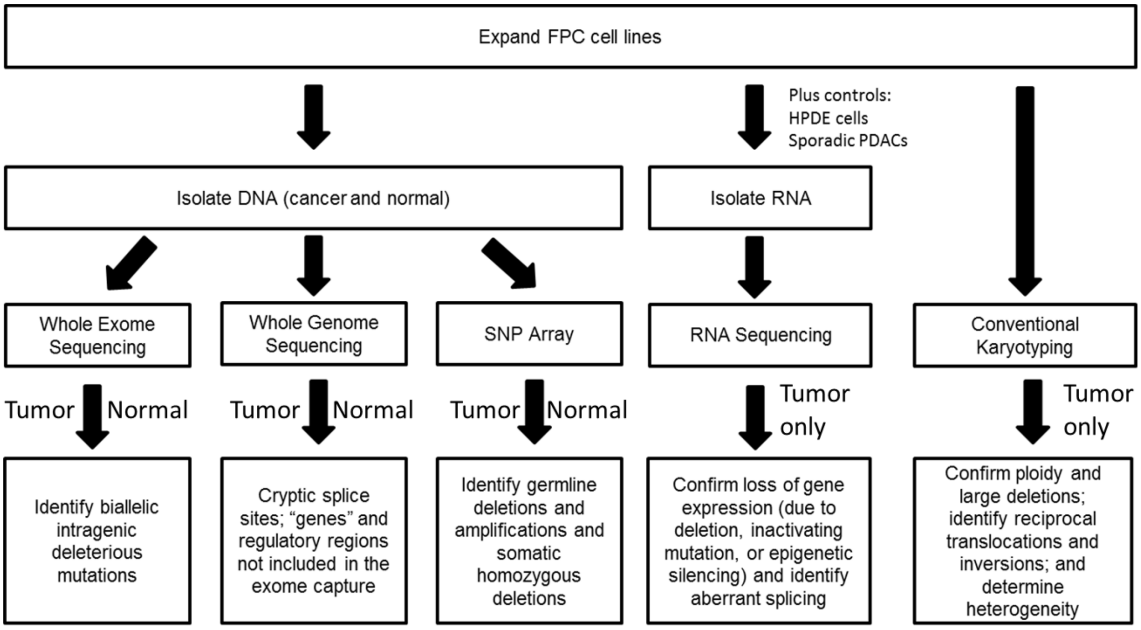


Figure 4.1. Integrative approach towards identifying new FPC predisposition genes. We have employed an integrated approach of analysis at the levels of chromosome (conventional karyotype and SNP array), DNA (whole exome and whole genome sequencing), and RNA (RNA sequencing).

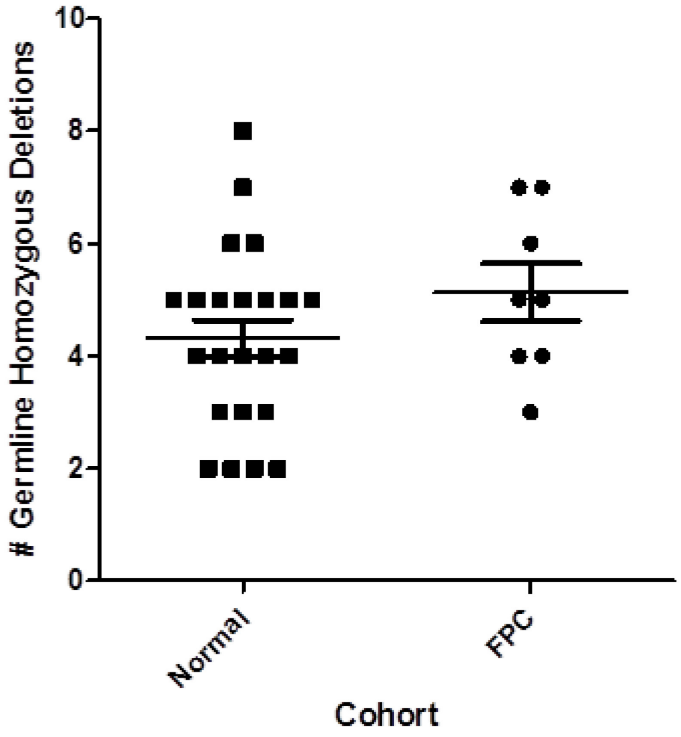


Figure 4.2. Germline intragenic HDs in normal and FPC cohorts. Normal samples include 23 normal individuals and FPC includes the 8 FPC patients.

Table 4.1. Value of an integrated approach

Alteration	Karyotype	SNP Microarray	WES	WGS	RNA-Seq
CN-LOH		X	+/-	+/-	+/-
Reciprocal Translocation	X		+/-	X	+/-
Inversion	+/-			X	+/-
Large Deletion	+/-	X	+/-	X	
Deleterious Intragenic Mutation			X	X	X
Small Intragenic Deletion		+/-	+/-	X	X
Aberrant Splicing					X
Epigenetic Silencing (e.g. methylation)					X

X=should detect; +/-=may or may not detect, copy-neutral loss of heterozygosity (CN-LOH), whole exome sequencing (WES), whole genome sequencing (WGS)

Table 4.2. Germline intragenic HDs in FPC samples and normal controls (CIDR).

Sample	#HDs	Total DNA Deleted (bp)
Pa101N	4	22,903
Pa102N	4	20,368
Pa222N	5	64,252
Pa227N	7	66,552
Pa228N	5	178,822
Pa229N	3	246,738
Pa230N	7	69,076
Pa231N	6	269,578
FPC_Med	**5**	**67,814**
FPC_Avg	**5.1**	**117,286**
CIDR_06985	5	180,569
CIDR_07048	5	95,579
CIDR_10846	8	557,405
CIDR_10847	2	48,940
CIDR_10851	4	69,217
CIDR_10860	3	84,972
CIDR_10861	4	745,759
CIDR_11881	6	93,058
CIDR_11882	5	30,607
CIDR_11992	5	242,264
CIDR_11993	4	89,731
CIDR_11994	4	95,599
CIDR_11995	3	9,963
CIDR_12056	2	31,184
CIDR_12057	7	123,722
CIDR_12144	5	166,924
CIDR_12145	3	13,823
CIDR_12146	5	24,063
CIDR_12155	2	76,359
CIDR_12156	4	194,081
CIDR_12239	5	184,298
CIDR_18501	2	7,930
CIDR_18859	6	84,293
CIDR_Med	**4**	**89,731**
CIDR_Avg	**4.3**	**141,319**
***p*-value** (unpaired, two-tailed t-test)	**0.204**	**0.641**

Chapter 5

Significance and future directions

Biological Basis for the Similar Age of Onset in Familial and Sporadic PDAC (Chapter 2)

We have shown that FPC and sporadic PDAC have a similar age of onset. It remains a conundrum in the field as to how one can inherit a predisposition to cancer without accelerating the age of onset. We propose that it is a combination of shared driver genes and a pancreatic tissue-specific effect. This hypothesis represents an important area for future study as it may shed light on PDAC carcinogenesis in general. Is there a "5th gene", such as a "gatekeeper" gene that is age-independent but increases the frequency or likelihood of developing PDAC, such as immune dysfunction, known to decrease with age?

Elucidation of TransFlip Prevalence and Mechanism of Production (Chapter 3)

We have discovered a new mechanism of TSG inactivation, TransFlip mutations. In FPC, TSGs are more commonly deleted by TransFlip mutations than simple interstitial deletions. TransFlip mutations were also found in sporadic PDAC, and it remains to be seen what the prevalence of TransFlip mutations are across all cancers, and that knowledge may give insight to the mechanism that gives rise to TransFlip mutations. Are TransFlip mutations ever the cause of an inherited human genetic disease? Are TransFlip mutations the manifestation of a specific DNA repair defect? If so, then TransFlip mutations may confer sensitivity to drugs that inhibit parallel pathways, in a synthetic lethality mechanism similar to *BRCA2* mutations and PARP inhibitors.

Identification of All Genes that Predispose to FPC (Chapter 4)

Our effort toward identifying new FPC predisposition genes highlights the challenge of such an endeavor. Major challenges of filter-based gene identification are (1) underlying genetic heterogeneity of FPC, (2) many putative deleterious mutations and even homozygous deletions are common variants, seen in the germline of normal individuals, making it hard to distinguish tumorigenic mutations from benign or tolerated mutations in a background of 1000s of germline mutations in a given sample, and (3) phenocopies, where patients in a family with FPC do not carry the genetic predisposition, but instead are a sporadic PDAC case.

An additional challenge is that yet-discovered predisposition genes may promote tumorigenesis in ways that are atypical. This includes haploinsufficiency, dominant negative mutations, and intergenic (non-coding) mutations that disrupt gene regulation. It remains to be seen if the lack of an obvious candidate predisposition gene is due to (1) the methods employed in our study are unable to detect the mechanism of gene inactivation (e.g. a post-translational modification that would require proteomics analysis), (2) inadequate tools to analyze the data (e.g. annotation issues that may preclude identification of coding mutations), or (3) shortcomings of knowledge of the human genome complexity (e.g. regulatory regions that are not currently annotated). While our analysis focused on intragenic mutations in protein-coding genes, as the understanding of the human genome improves, our current data can be re-analyzed for newly identified regulatory mutations in non-coding regions and expression levels of non-coding RNAs.

Impact on Cancer Patients

In summary, our results indicate the similarity of FPC to its sporadic counterpart, with respect to age of onset and genetic characteristics. This highlights the utility of early detection efforts tested in familial cohorts to the larger population. However, early detection testing that utilizes PARE will require more sensitive and specific SV calling algorithms.

We have discovered a new class of SV, TransFlip mutations, which are present in both familial and sporadic PDAC. The lack of TransFlip mutations in the germline suggests that they are completely unique to cancer. The most exciting implication of TransFlip mutations is that they are potentially targetable via synthetic lethality, similar to *BRCA* mutations conferring sensitivity to poly ADP ribose polymerase (PARP) inhibitors or by exploiting genes in the HD, like in the co-deletion of *MTAP* with *p16* (*79, 80*).

References

1. N. A. Howlader N, Krapcho M, Garshell J, Neyman N, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Cho H, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds), SEER Cancer Statistics Review, 1975-2010, National Cancer Institute. *SEER web site*, (April 2013).
2. R. H. Hruban, R. E. Wilentz, S. E. Kern, Genetic progression in the pancreatic ducts. *The American journal of pathology* **156**, 1821-1825 (2000); published online EpubJun (10.1016/S0002-9440(10)65054-7).
3. G. M. Petersen, R. H. Hruban, Familial pancreatic cancer: where are we in 2003? *Journal of the National Cancer Institute* **95**, 180-181 (2003); published online EpubFeb 5 (
4. C. Shi, R. H. Hruban, A. P. Klein, Familial pancreatic cancer. *Archives of pathology & laboratory medicine* **133**, 365-374 (2009); published online EpubMar (10.1043/1543-2165-133.3.365).
5. D. J. Bunyan, J. Shea-Simonds, A. C. Reck, D. Finnis, D. M. Eccles, Genotype-phenotype correlations of new causative APC gene mutations in patients with familial adenomatous polyposis. *Journal of medical genetics* **32**, 728-731 (1995); published online EpubSep (
6. A. M. Goldstein, M. Chan, M. Harland, E. M. Gillanders, N. K. Hayward, M. F. Avril, E. Azizi, G. Bianchi-Scarra, D. T. Bishop, B. Bressac-de Paillerets, W. Bruno, D. Calista, L. A. Cannon Albright, F. Demenais, D. E. Elder, P. Ghiorzo, N. A. Gruis, J. Hansson, D. Hogg, E. A. Holland, P. A. Kanetsky, R. F. Kefford, M. T. Landi, J. Lang, S. A. Leachman, R. M. Mackie, V. Magnusson, G. J. Mann, K. Niendorf, J. Newton Bishop, J. M. Palmer, S. Puig, J. A. Puig-Butille, F. A. de Snoo, M. Stark, H. Tsao, M. A. Tucker, L. Whitaker, E. Yakobson, C. Melanoma Genetics, High-risk melanoma susceptibility genes and pancreatic cancer, neural system tumors, and uveal melanoma across GenoMEL. *Cancer research* **66**, 9818-9828 (2006); published online EpubOct 15 (10.1158/0008-5472.CAN-06-0494).
7. J. M. Hall, M. K. Lee, B. Newman, J. E. Morrow, L. A. Anderson, B. Huey, M. C. King, Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**, 1684-1689 (1990); published online EpubDec 21 (
8. T. Myrhoj, M. L. Bisgaard, I. Bernstein, L. B. Svendsen, J. O. Sondergaard, S. Bulow, Hereditary non-polyposis colorectal cancer: clinical features and survival. Results from the Danish HNPCC register. *Scandinavian journal of gastroenterology* **32**, 572-576 (1997); published online EpubJun (
9. M. Krainer, S. Silva-Arrieta, M. G. FitzGerald, A. Shimada, C. Ishioka, R. Kanamaru, D. J. MacDonald, H. Unsal, D. M. Finkelstein, A. Bowcock, K. J. Isselbacher, D. A. Haber, Differential contributions of BRCA1 and BRCA2 to early-onset breast cancer. *The New England journal of medicine* **336**, 1416-1421 (1997); published online EpubMay 15 (10.1056/NEJM199705153362003).
10. S. Raimondi, P. Maisonneuve, J. M. Lohr, A. B. Lowenfels, Early onset pancreatic cancer: evidence of a major role for smoking and genetic factors. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **16**, 1894-1897 (2007); published online EpubSep (10.1158/1055-9965.EPI-07-0341).
11. S. Yachida, S. Jones, I. Bozic, T. Antal, R. Leary, B. Fu, M. Kamiyama, R. H. Hruban, J. R. Eshleman, M. A. Nowak, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, C. A. Iacobuzio-Donahue, Distant metastasis occurs late during the genetic evolution of pancreatic

cancer. *Nature* **467**, 1114-1117 (2010); published online EpubOct 28 (10.1038/nature09515).
12. J. R. Eshleman, A. L. Norris, Y. Sadakari, M. Debeljak, M. Borges, C. Harrington, E. Lin, A. Brant, T. Barkley, J. A. Almario, M. Topazian, J. Farrell, S. Syngal, J. H. Lee, J. Yu, R. H. Hruban, M. Kanda, M. I. Canto, M. Goggins, KRAS and GNAS Mutations in Pancreatic Juice Collected From the Duodenum of Patients at High Risk for Neoplasia Undergoing Endoscopic Ultrasound. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association*, (2014); published online EpubDec 3 (10.1016/j.cgh.2014.11.028).
13. R. J. Leary, I. Kinde, F. Diehl, K. Schmidt, C. Clouser, C. Duncan, A. Antipova, C. Lee, K. McKernan, F. M. De La Vega, K. W. Kinzler, B. Vogelstein, L. A. Diaz, Jr., V. E. Velculescu, Development of personalized tumor biomarkers using massively parallel sequencing. *Science translational medicine* **2**, 20ra14 (2010); published online EpubFeb 24 (10.1126/scitranslmed.3000702).
14. A. Haddad, G. C. Kowdley, T. M. Pawlik, S. C. Cunningham, Hereditary pancreatic and hepatobiliary cancers. *International journal of surgical oncology* **2011**, 154673 (2011)10.1155/2011/154673).
15. R. H. Hruban, M. I. Canto, M. Goggins, R. Schulick, A. P. Klein, Update on familial pancreatic cancer. *Advances in surgery* **44**, 293-311 (2010).
16. A. P. Klein, K. A. Brune, G. M. Petersen, M. Goggins, A. C. Tersmette, G. J. Offerhaus, C. Griffin, J. L. Cameron, C. J. Yeo, S. Kern, R. H. Hruban, Prospective risk of pancreatic cancer in familial pancreatic cancer kindreds. *Cancer research* **64**, 2634-2638 (2004); published online EpubApr 1 (
17. N. J. Roberts, Y. Jiao, J. Yu, L. Kopelovich, G. M. Petersen, M. L. Bondy, S. Gallinger, A. G. Schwartz, S. Syngal, M. L. Cote, J. Axilbund, R. Schulick, S. Z. Ali, J. R. Eshleman, V. E. Velculescu, M. Goggins, B. Vogelstein, N. Papadopoulos, R. H. Hruban, K. W. Kinzler, A. P. Klein, ATM mutations in patients with hereditary pancreatic cancer. *Cancer discovery* **2**, 41-46 (2012); published online EpubJan (10.1158/2159-8290.CD-11-0194).
18. C. Caldas, S. A. Hahn, L. T. da Costa, M. S. Redston, M. Schutte, A. B. Seymour, C. L. Weinstein, R. H. Hruban, C. J. Yeo, S. E. Kern, Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma. *Nature genetics* **8**, 27-32 (1994); published online EpubSep (10.1038/ng0994-27).
19. C. A. Iacobuzio-Donahue, M. S. van der Heijden, M. R. Baumgartner, W. J. Troup, J. M. Romm, K. Doheny, E. Pugh, C. J. Yeo, M. G. Goggins, R. H. Hruban, S. E. Kern, Large-scale allelotype of pancreaticobiliary carcinoma provides quantitative estimates of genome-wide allelic loss. *Cancer research* **64**, 871-875 (2004); published online EpubFeb 1 (
20. S. Jones, X. Zhang, D. W. Parsons, J. C. Lin, R. J. Leary, P. Angenendt, P. Mankoo, H. Carter, H. Kamiyama, A. Jimeno, S. M. Hong, B. Fu, M. T. Lin, E. S. Calhoun, M. Kamiyama, K. Walter, T. Nikolskaya, Y. Nikolsky, J. Hartigan, D. R. Smith, M. Hidalgo, S. D. Leach, A. P. Klein, E. M. Jaffee, M. Goggins, A. Maitra, C. Iacobuzio-Donahue, J. R. Eshleman, S. E. Kern, R. H. Hruban, R. Karchin, N. Papadopoulos, G. Parmigiani, B. Vogelstein, V. E. Velculescu, K. W. Kinzler, Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321**, 1801-1806 (2008); published online EpubSep 26 (10.1126/science.1164368).
21. A. V. Biankin, N. Waddell, K. S. Kassahn, M. C. Gingras, L. B. Muthuswamy, A. L. Johns, D. K. Miller, P. J. Wilson, A. M. Patch, J. Wu, D. K. Chang, M. J. Cowley, B. B. Gardiner, S. Song, I. Harliwong, S. Idrisoglu, C. Nourse, E. Nourbakhsh, S. Manning, S. Wani, M. Gongora, M. Pajic, C. J. Scarlett, A. J. Gill, A. V. Pinho, I. Rooman, M. Anderson, O.

Holmes, C. Leonard, D. Taylor, S. Wood, Q. Xu, K. Nones, J. L. Fink, A. Christ, T. Bruxner, N. Cloonan, G. Kolle, F. Newell, M. Pinese, R. S. Mead, J. L. Humphris, W. Kaplan, M. D. Jones, E. K. Colvin, A. M. Nagrial, E. S. Humphrey, A. Chou, V. T. Chin, L. A. Chantrill, A. Mawson, J. S. Samra, J. G. Kench, J. A. Lovell, R. J. Daly, N. D. Merrett, C. Toon, K. Epari, N. Q. Nguyen, A. Barbour, N. Zeps, I. Australian Pancreatic Cancer Genome, N. Kakkar, F. Zhao, Y. Q. Wu, M. Wang, D. M. Muzny, W. E. Fisher, F. C. Brunicardi, S. E. Hodges, J. G. Reid, J. Drummond, K. Chang, Y. Han, L. R. Lewis, H. Dinh, C. J. Buhay, T. Beck, L. Timms, M. Sam, K. Begley, A. Brown, D. Pai, A. Panchal, N. Buchner, R. De Borja, R. E. Denroche, C. K. Yung, S. Serra, N. Onetto, D. Mukhopadhyay, M. S. Tsao, P. A. Shaw, G. M. Petersen, S. Gallinger, R. H. Hruban, A. Maitra, C. A. Iacobuzio-Donahue, R. D. Schulick, C. L. Wolfgang, R. A. Morgan, R. T. Lawlor, P. Capelli, V. Corbo, M. Scardoni, G. Tortora, M. A. Tempero, K. M. Mann, N. A. Jenkins, P. A. Perez-Mancera, D. J. Adams, D. A. Largaespada, L. F. Wessels, A. G. Rust, L. D. Stein, D. A. Tuveson, N. G. Copeland, E. A. Musgrove, A. Scarpa, J. R. Eshleman, T. J. Hudson, R. L. Sutherland, D. A. Wheeler, J. V. Pearson, J. D. McPherson, R. A. Gibbs, S. M. Grimmond, Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* **491**, 399-405 (2012); published online EpubNov 15 (10.1038/nature11547).

22. K. Brune, T. Abe, M. Canto, L. O'Malley, A. P. Klein, A. Maitra, N. Volkan Adsay, E. K. Fishman, J. L. Cameron, C. J. Yeo, S. E. Kern, M. Goggins, R. H. Hruban, Multifocal neoplastic precursor lesions associated with lobular atrophy of the pancreas in patients having a strong family history of pancreatic cancer. *The American journal of surgical pathology* **30**, 1067-1076 (2006); published online EpubSep (pas.0000213265.84725.0b).
23. C. Shi, A. P. Klein, M. Goggins, A. Maitra, M. Canto, S. Ali, R. Schulick, E. Palmisano, R. H. Hruban, Increased Prevalence of Precursor Lesions in Familial Pancreatic Cancer Patients. *Clinical cancer research : an official journal of the American Association for Cancer Research* **15**, 7737-7743 (2009); published online EpubDec 15 (10.1158/1078-0432.CCR-09-0004).
24. A. D. Singhi, H. Ishida, S. Ali, M. Goggins, M. Canto, C. L. Wolfgang, Z. Meriden, N. J. Roberts, A. P. Klein, R. H. Hruban, A histomorphologic comparison of familial and sporadic pancreatic cancers. (Submitted).
25. C. Bettegowda, M. Sausen, R. J. Leary, I. Kinde, Y. Wang, N. Agrawal, B. R. Bartlett, H. Wang, B. Luber, R. M. Alani, E. S. Antonarakis, N. S. Azad, A. Bardelli, H. Brem, J. L. Cameron, C. C. Lee, L. A. Fecher, G. L. Gallia, P. Gibbs, D. Le, R. L. Giuntoli, M. Goggins, M. D. Hogarty, M. Holdhoff, S. M. Hong, Y. Jiao, H. H. Juhl, J. J. Kim, G. Siravegna, D. A. Laheru, C. Lauricella, M. Lim, E. J. Lipson, S. K. Marie, G. J. Netto, K. S. Oliner, A. Olivi, L. Olsson, G. J. Riggins, A. Sartore-Bianchi, K. Schmidt, M. Shih l, S. M. Oba-Shinjo, S. Siena, D. Theodorescu, J. Tie, T. T. Harkins, S. Veronese, T. L. Wang, J. D. Weingart, C. L. Wolfgang, L. D. Wood, D. Xing, R. H. Hruban, J. Wu, P. J. Allen, C. M. Schmidt, M. A. Choti, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, N. Papadopoulos, L. A. Diaz, Jr., Detection of circulating tumor DNA in early- and late-stage human malignancies. *Science translational medicine* **6**, 224ra224 (2014); published online EpubFeb 19 (10.1126/scitranslmed.3007094).
26. E. M. Nicholls, Somatic variation and multiple neurofibromatosis. *Human heredity* **19**, 473-479 (1969).
27. A. G. Knudson, Jr., Mutation and cancer: statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America* **68**, 820-823 (1971); published online EpubApr (

28. C. A. Griffin, L. Morsberger, A. L. Hawkins, M. Haddadin, A. Patel, T. Ried, E. Schrock, E. J. Perlman, E. Jaffee, Molecular cytogenetic characterization of pancreas cancer cell lines reveals high complexity chromosomal alterations. *Cytogenetic and genome research* **118**, 148-156 (2007)10.1159/000108295).
29. H. Ouyang, L. Mou, C. Luk, N. Liu, J. Karaskova, J. Squire, M. S. Tsao, Immortal human pancreatic duct epithelial cell lines with near normal genotype and phenotype. *The American journal of pathology* **157**, 1623-1631 (2000); published online EpubNov (10.1016/S0002-9440(10)64800-6).
30. Y. Jiao, C. Shi, B. H. Edil, R. F. de Wilde, D. S. Klimstra, A. Maitra, R. D. Schulick, L. H. Tang, C. L. Wolfgang, M. A. Choti, V. E. Velculescu, L. A. Diaz, Jr., B. Vogelstein, K. W. Kinzler, R. H. Hruban, N. Papadopoulos, DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science* **331**, 1199-1203 (2011); published online EpubMar 4 (10.1126/science.1200609).
31. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* **12**, 323 (2011)10.1186/1471-2105-12-323).
32. H. Thorvaldsdottir, J. T. Robinson, J. P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics* **14**, 178-192 (2013); published online EpubMar (10.1093/bib/bbs017).
33. N. C. I. h. o. t. Internet], Surveillance Epidemiology and End Results (SEER) Program. SEER*Stat Database: Incidence - SEER 13 Regs Public-Use, Nov 2004 Sub (1973–2002 varying), National Cancer Institute DCCPS, Surveillance Research Program, Cancer Statistics Branch. http://www.seer.cancer.gov., (released April 2005, based on the November 2004 submission [cited 2014 Jun 24]).
34. K. Brune, S. M. Hong, A. Li, S. Yachida, T. Abe, M. Griffith, D. Yang, N. Omura, J. Eshleman, M. Canto, R. Schulick, A. P. Klein, R. H. Hruban, C. Iacobuzio-Donohue, M. Goggins, Genetic and epigenetic alterations of familial pancreatic cancers. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **17**, 3536-3542 (2008); published online EpubDec (10.1158/1055-9965.EPI-08-0630).
35. Y. Jiao, R. Yonescu, G. J. Offerhaus, D. S. Klimstra, A. Maitra, J. R. Eshleman, J. G. Herman, W. Poh, L. Pelosof, C. L. Wolfgang, B. Vogelstein, K. W. Kinzler, R. H. Hruban, N. Papadopoulos, L. D. Wood, Whole-exome sequencing of pancreatic neoplasms with acinar differentiation. *The Journal of pathology* **232**, 428-435 (2014); published online EpubMar (10.1002/path.4310).
36. J. Wu, Y. Jiao, M. Dal Molin, A. Maitra, R. F. de Wilde, L. D. Wood, J. R. Eshleman, M. G. Goggins, C. L. Wolfgang, M. I. Canto, R. D. Schulick, B. H. Edil, M. A. Choti, V. Adsay, D. S. Klimstra, G. J. Offerhaus, A. P. Klein, L. Kopelovich, H. Carter, R. Karchin, P. J. Allen, C. M. Schmidt, Y. Naito, L. A. Diaz, Jr., K. W. Kinzler, N. Papadopoulos, R. H. Hruban, B. Vogelstein, Whole-exome sequencing of neoplastic cysts of the pancreas reveals recurrent mutations in components of ubiquitin-dependent pathways. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 21188-21193 (2011); published online EpubDec 27 (10.1073/pnas.1118046108).
37. J. Wu, H. Matthaei, A. Maitra, M. Dal Molin, L. D. Wood, J. R. Eshleman, M. Goggins, M. I. Canto, R. D. Schulick, B. H. Edil, C. L. Wolfgang, A. P. Klein, L. A. Diaz, Jr., P. J. Allen, C. M. Schmidt, K. W. Kinzler, N. Papadopoulos, R. H. Hruban, B. Vogelstein, Recurrent GNAS mutations define an unexpected pathway for pancreatic cyst development.

Science translational medicine **3**, 92ra66 (2011); published online EpubJul 20 (10.1126/scitranslmed.3002543).
38. J. Luttges, C. Stigge, M. Pacena, G. Kloppel, Rare ductal adenocarcinoma of the pancreas in patients younger than age 40 years. *Cancer* **100**, 173-182 (2004); published online EpubJan 1 (10.1002/cncr.11860).
39. G. Lal, G. Liu, B. Schmocker, P. Kaurah, H. Ozcelik, S. A. Narod, M. Redston, S. Gallinger, Inherited predisposition to pancreatic adenocarcinoma: role of family history and germ-line p16, BRCA1, and BRCA2 mutations. *Cancer research* **60**, 409-416 (2000); published online EpubJan 15 (
40. T. A. James, D. G. Sheldon, A. Rajput, B. W. Kuvshinoff, M. M. Javle, H. R. Nava, J. L. Smith, J. F. Gibbs, Risk factors associated with earlier age of onset in familial pancreatic carcinoma. *Cancer* **101**, 2722-2726 (2004); published online EpubDec 15 (10.1002/cncr.20700).
41. A. Brandt, J. L. Bermejo, J. Sundquist, K. Hemminki, Age of onset in familial cancer. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* **19**, 2084-2088 (2008); published online EpubDec (10.1093/annonc/mdn527).
42. G. M. Petersen, M. de Andrade, M. Goggins, R. H. Hruban, M. Bondy, J. F. Korczak, S. Gallinger, H. T. Lynch, S. Syngal, K. G. Rabe, D. Seminara, A. P. Klein, Pancreatic cancer genetic epidemiology consortium. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **15**, 704-710 (2006); published online EpubApr (10.1158/1055-9965.EPI-05-0734).
43. M. A. Eberle, R. Pfutzer, K. L. Pogue-Geile, M. P. Bronner, D. Crispin, M. B. Kimmey, R. H. Duerr, L. Kruglyak, D. C. Whitcomb, T. A. Brentnall, A new susceptibility locus for autosomal dominant pancreatic cancer maps to chromosome 4q32-34. *American journal of human genetics* **70**, 1044-1048 (2002); published online EpubApr (10.1086/339692).
44. H. Ozcelik, B. Schmocker, N. Di Nicola, X. H. Shi, B. Langer, M. Moore, B. R. Taylor, S. A. Narod, G. Darlington, I. L. Andrulis, S. Gallinger, M. Redston, Germline BRCA2 6174delT mutations in Ashkenazi Jewish pancreatic cancer patients. *Nature genetics* **16**, 17-18 (1997); published online EpubMay (10.1038/ng0597-17).
45. E. Fernandez, C. La Vecchia, B. D'Avanzo, E. Negri, S. Franceschi, Family history and the risk of liver, gallbladder, and pancreatic cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **3**, 209-212 (1994); published online EpubApr-May (
46. R. R. McWilliams, W. R. Bamlet, K. G. Rabe, J. E. Olson, M. de Andrade, G. M. Petersen, Association of family history of specific cancers with a younger age of onset of pancreatic adenocarcinoma. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association* **4**, 1143-1147 (2006); published online EpubSep (10.1016/j.cgh.2006.05.029).
47. D. K. Bartsch, R. Kress, M. Sina-Frey, R. Grutzmann, B. Gerdes, C. Pilarsky, J. W. Heise, K. M. Schulte, M. Colombo-Benkmann, C. Schleicher, H. Witzigmann, O. Pridohl, M. B. Ghadimi, O. Horstmann, W. von Bernstorff, L. Jochimsen, J. Schmidt, S. Eisold, L. Estevez-Schwarz, S. A. Hahn, K. Schulmann, W. Bock, T. M. Gress, N. Zugel, K. Breitschaft, K. Prenzel, H. Messmann, E. Endlicher, M. Schneider, A. Ziegler, W. Schmiegel, H. Schafer, M. Rothmund, H. Rieder, Prevalence of familial pancreatic cancer

in Germany. *International journal of cancer. Journal international du cancer* **110**, 902-906 (2004); published online EpubJul 20 (10.1002/ijc.20210).
48. H. T. Lynch, M. L. Fitzsimmons, T. C. Smyrk, S. J. Lanspa, P. Watson, J. McClellan, J. F. Lynch, Familial pancreatic cancer: clinicopathologic study of 18 nuclear families. *The American journal of gastroenterology* **85**, 54-60 (1990); published online EpubJan (
49. H. T. Lynch, L. Fusaro, J. F. Lynch, Familial pancreatic cancer: a family study. *Pancreas* **7**, 511-515 (1992).
50. J. Ji, A. Forsti, J. Sundquist, P. Lenner, K. Hemminki, Survival in familial pancreatic cancer. *Pancreatology : official journal of the International Association of Pancreatology* **8**, 252-256 (2008)10.1159/000134272).
51. P. Ghadirian, P. Boyle, A. Simard, J. Baillargeon, P. Maisonneuve, C. Perret, Reported family aggregation of pancreatic cancer within a population-based case-control study in the Francophone community in Montreal, Canada. *International journal of pancreatology : official journal of the International Association of Pancreatology* **10**, 183-196 (1991); published online EpubNov-Dec (
52. R. H. Hruban, G. M. Petersen, P. K. Ha, S. E. Kern, Genetics of pancreatic cancer. From genes to families. *Surgical oncology clinics of North America* **7**, 1-23 (1998); published online EpubJan (
53. D. T. Silverman, M. Schiffman, J. Everhart, A. Goldstein, K. D. Lillemoe, G. M. Swanson, A. G. Schwartz, L. M. Brown, R. S. Greenberg, J. B. Schoenberg, L. M. Pottern, R. N. Hoover, J. F. Fraumeni, Jr., Diabetes mellitus, other medical conditions and familial history of cancer as risk factors for pancreatic cancer. *British journal of cancer* **80**, 1830-1837 (1999); published online EpubAug (10.1038/sj.bjc.6690607).
54. J. Permuth-Wey, K. M. Egan, Family history is a significant risk factor for pancreatic cancer: results from a systematic review and meta-analysis. *Familial cancer* **8**, 109-117 (2009)10.1007/s10689-008-9214-8).
55. F. J. Couch, M. R. Johnson, K. G. Rabe, K. Brune, M. de Andrade, M. Goggins, H. Rothenmund, S. Gallinger, A. Klein, G. M. Petersen, R. H. Hruban, The prevalence of BRCA2 mutations in familial pancreatic cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **16**, 342-346 (2007); published online EpubFeb (10.1158/1055-9965.EPI-06-0783).
56. N. Al-Tassan, N. H. Chmiel, J. Maynard, N. Fleming, A. L. Livingston, G. T. Williams, A. K. Hodges, D. R. Davies, S. S. David, J. R. Sampson, J. P. Cheadle, Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. *Nature genetics* **30**, 227-232 (2002); published online EpubFeb (10.1038/ng828).
57. M. G. FitzGerald, D. P. Harkin, S. Silva-Arrieta, D. J. MacDonald, L. C. Lucchina, H. Unsal, E. O'Neill, J. Koh, D. M. Finkelstein, K. J. Isselbacher, A. J. Sober, D. A. Haber, Prevalence of germ-line mutations in p16, p19ARF, and CDK4 in familial melanoma: analysis of a clinic-based population. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 8541-8545 (1996); published online EpubAug 6 (
58. A. M. Goldstein, J. P. Struewing, A. Chidambaram, M. C. Fraser, M. A. Tucker, Genotype-phenotype relationships in U.S. melanoma-prone families with CDKN2A and CDK4 mutations. *Journal of the National Cancer Institute* **92**, 1006-1010 (2000); published online EpubJun 21 (
59. S. A. Hahn, M. Schutte, A. T. Hoque, C. A. Moskaluk, L. T. da Costa, E. Rozenblum, C. L. Weinstein, A. Fischer, C. J. Yeo, R. H. Hruban, S. E. Kern, DPC4, a candidate tumor

suppressor gene at human chromosome 18q21.1. *Science* **271**, 350-353 (1996); published online EpubJan 19 (

60. W. H. Lee, R. Bookstein, F. Hong, L. J. Young, J. Y. Shew, E. Y. Lee, Human retinoblastoma susceptibility gene: cloning, identification, and sequence. *Science* **235**, 1394-1399 (1987); published online EpubMar 13 (
61. G. R. Bignell, C. D. Greenman, H. Davies, A. P. Butler, S. Edkins, J. M. Andrews, G. Buck, L. Chen, D. Beare, C. Latimer, S. Widaa, J. Hinton, C. Fahey, B. Fu, S. Swamy, G. L. Dalgliesh, B. T. Teh, P. Deloukas, F. Yang, P. J. Campbell, P. A. Futreal, M. R. Stratton, Signatures of mutation and selection in the cancer genome. *Nature* **463**, 893-898 (2010); published online EpubFeb 18 (10.1038/nature08768).
62. P. J. Stephens, C. D. Greenman, B. Fu, F. Yang, G. R. Bignell, L. J. Mudie, E. D. Pleasance, K. W. Lau, D. Beare, L. A. Stebbings, S. McLaren, M. L. Lin, D. J. McBride, I. Varela, S. Nik-Zainal, C. Leroy, M. Jia, A. Menzies, A. P. Butler, J. W. Teague, M. A. Quail, J. Burton, H. Swerdlow, N. P. Carter, L. A. Morsberger, C. Iacobuzio-Donahue, G. A. Follows, A. R. Green, A. M. Flanagan, M. R. Stratton, P. A. Futreal, P. J. Campbell, Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27-40 (2011); published online EpubJan 7 (10.1016/j.cell.2010.11.055).
63. C. Chiang, J. C. Jacobsen, C. Ernst, C. Hanscom, A. Heilbut, I. Blumenthal, R. E. Mills, A. Kirby, A. M. Lindgren, S. R. Rudiger, C. J. McLaughlan, C. S. Bawden, S. J. Reid, R. L. Faull, R. G. Snell, I. M. Hall, Y. Shen, T. K. Ohsumi, M. L. Borowsky, M. J. Daly, C. Lee, C. C. Morton, M. E. MacDonald, J. F. Gusella, M. E. Talkowski, Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration. *Nature genetics* **44**, 390-397, S391 (2012); published online EpubApr (10.1038/ng.2202).
64. S. C. Baca, D. Prandi, M. S. Lawrence, J. M. Mosquera, A. Romanel, Y. Drier, K. Park, N. Kitabayashi, T. Y. MacDonald, M. Ghandi, E. Van Allen, G. V. Kryukov, A. Sboner, J. P. Theurillat, T. D. Soong, E. Nickerson, D. Auclair, A. Tewari, H. Beltran, R. C. Onofrio, G. Boysen, C. Guiducci, C. E. Barbieri, K. Cibulskis, A. Sivachenko, S. L. Carter, G. Saksena, D. Voet, A. H. Ramos, W. Winckler, M. Cipicchio, K. Ardlie, P. W. Kantoff, M. F. Berger, S. B. Gabriel, T. R. Golub, M. Meyerson, E. S. Lander, O. Elemento, G. Getz, F. Demichelis, M. A. Rubin, L. A. Garraway, Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666-677 (2013); published online EpubApr 25 (10.1016/j.cell.2013.03.021).
65. P. J. Campbell, S. Yachida, L. J. Mudie, P. J. Stephens, E. D. Pleasance, L. A. Stebbings, L. A. Morsberger, C. Latimer, S. McLaren, M. L. Lin, D. J. McBride, I. Varela, S. A. Nik-Zainal, C. Leroy, M. Jia, A. Menzies, A. P. Butler, J. W. Teague, C. A. Griffin, J. Burton, H. Swerdlow, M. A. Quail, M. R. Stratton, C. Iacobuzio-Donahue, P. A. Futreal, The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467**, 1109-1113 (2010); published online EpubOct 28 (10.1038/nature09460).
66. J. M. Tubio, Y. Li, Y. S. Ju, I. Martincorena, S. L. Cooke, M. Tojo, G. Gundem, C. P. Pipinikas, J. Zamora, K. Raine, A. Menzies, P. Roman-Garcia, A. Fullam, M. Gerstung, A. Shlien, P. S. Tarpey, E. Papaemmanuil, S. Knappskog, P. Van Loo, M. Ramakrishna, H. R. Davies, J. Marshall, D. C. Wedge, J. W. Teague, A. P. Butler, S. Nik-Zainal, L. Alexandrov, S. Behjati, L. R. Yates, N. Bolli, L. Mudie, C. Hardy, S. Martin, S. McLaren, S. O'Meara, E. Anderson, M. Maddison, S. Gamble, I. B. C. Group, I. B. C. Group, I. P. C. Group, C. Foster, A. Y. Warren, H. Whitaker, D. Brewer, R. Eeles, C. Cooper, D. Neal, A. G. Lynch, T. Visakorpi, W. B. Isaacs, L. van't Veer, C. Caldas, C. Desmedt, C. Sotiriou, S. Aparicio, J. A. Foekens, J. E. Eyfjord, S. R. Lakhani, G. Thomas, O. Myklebost, P. N. Span, A. L. Borresen-Dale, A. L. Richardson, M. Van de Vijver, A. Vincent-Salomon, G. G. Van den Eynden, A.

M. Flanagan, P. A. Futreal, S. M. Janes, G. S. Bova, M. R. Stratton, U. McDermott, P. J. Campbell, Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* **345**, 1251343 (2014); published online EpubAug 1 (10.1126/science.1251343).

67. C. Cox, G. Bignell, C. Greenman, A. Stabenau, W. Warren, P. Stephens, H. Davies, S. Watt, J. Teague, S. Edkins, E. Birney, D. F. Easton, R. Wooster, P. A. Futreal, M. R. Stratton, A survey of homozygous deletions in human cancer genomes. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 4542-4547 (2005); published online EpubMar 22 (10.1073/pnas.0408593102).
68. A. L. Norris, N. J. Roberts, S. Jones, S. J. Wheelan, N. Papadopoulos, B. Vogelstein, K. W. Kinzler, R. H. Hruban, A. P. Klein, J. R. Eshleman, Familial and sporadic pancreatic cancer share the same molecular pathogenesis. *Familial cancer*, (2014); published online EpubSep 21 (10.1007/s10689-014-9755-y).
69. C. A. Griffin, R. H. Hruban, P. P. Long, L. A. Morsberger, F. Douna-Issa, C. J. Yeo, Chromosome abnormalities in pancreatic adenocarcinoma. *Genes, chromosomes & cancer* **9**, 93-100 (1994); published online EpubFeb (
70. E. E. Embuscado, D. Laheru, F. Ricci, K. J. Yun, S. de Boom Witzel, A. Seigel, K. Flickinger, M. Hidalgo, G. S. Bova, C. A. Iacobuzio-Donahue, Immortalizing the complexity of cancer metastasis: genetic features of lethal metastatic pancreatic cancer obtained from rapid autopsy. *Cancer biology & therapy* **4**, 548-554 (2005); published online EpubMay (
71. A. Untergasser, H. Nijveen, X. Rao, T. Bisseling, R. Geurts, J. A. Leunissen, Primer3Plus, an enhanced web interface to Primer3. *Nucleic acids research* **35**, W71-74 (2007); published online EpubJul (10.1093/nar/gkm306).
72. J. B. a. M. F. Mitelman F, Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer. *http://cgap.nci.nih.gov/Chromosomes/Mitelman*, (2015).
73. H. Ali, A. Daser, P. Dear, H. Wood, P. Rabbitts, T. Rabbitts, Nonreciprocal chromosomal translocations in renal cancer involve multiple DSBs and NHEJ associated with breakpoint inversion but not necessarily with transcription. *Genes, chromosomes & cancer* **52**, 402-409 (2013); published online EpubApr (10.1002/gcc.22038).
74. H. Herholz, W. Kern, S. Schnittger, T. Haferlach, F. Dicker, C. Haferlach, Translocations as a mechanism for homozygous deletion of 13q14 and loss of the ATM gene in a patient with B-cell chronic lymphocytic leukemia. *Cancer genetics and cytogenetics* **174**, 57-60 (2007); published online EpubApr 1 (10.1016/j.cancergencyto.2006.11.006).
75. A. Misawa, H. Hosoi, I. Imoto, T. Iehara, T. Sugimoto, J. Inazawa, Translocation (1;22)(p36;q11.2) with concurrent del(22)(q11.2) resulted in homozygous deletion of SNF5/INI1 in a newly established cell line derived from extrarenal rhabdoid tumor. *Journal of human genetics* **49**, 586-589 (2004)10.1007/s10038-004-0191-y).
76. J. M. Cayuela, B. Gardie, F. Sigaux, Disruption of the multiple tumor suppressor gene MTS1/p16(INK4a)/CDKN2 by illegitimate V(D)J recombinase activity in T-cell acute lymphoblastic leukemias. *Blood* **90**, 3720-3726 (1997); published online EpubNov 1 (
77. S. Sasaki, Y. Kitagawa, Y. Sekido, J. D. Minna, H. Kuwano, J. Yokota, T. Kohno, Molecular processes of chromosome 9p21 deletions in human cancers. *Oncogene* **22**, 3792-3798 (2003); published online EpubJun 12 (10.1038/sj.onc.1206589).
78. S. Raschke, V. Balz, T. Efferth, W. A. Schulz, A. R. Florl, Homozygous deletions of CDKN2A caused by alternative mechanisms in various human cancer cell lines. *Genes, chromosomes & cancer* **42**, 58-67 (2005); published online EpubJan (10.1002/gcc.20119).

79. H. Farmer, N. McCabe, C. J. Lord, A. N. Tutt, D. A. Johnson, T. B. Richardson, M. Santarosa, K. J. Dillon, I. Hickson, C. Knights, N. M. Martin, S. P. Jackson, G. C. Smith, A. Ashworth, Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917-921 (2005); published online EpubApr 14 (10.1038/nature03445).
80. S. R. Hustinx, R. H. Hruban, L. M. Leoni, C. Iacobuzio-Donahue, J. L. Cameron, C. J. Yeo, P. N. Brown, P. Argani, R. Ashfaq, N. Fukushima, M. Goggins, S. E. Kern, A. Maitra, Homozygous deletion of the MTAP gene in invasive adenocarcinoma of the pancreas and in periampullary cancer: a potential new target for therapy. *Cancer biology & therapy* **4**, 83-86 (2005); published online EpubJan (
81. T. Boveri, Zellenstudien. VI. Eine für die erste Orientierung geeignete Darstellung dieser und anderer Chromosomenprobleme findet sich in meiner Schrift: Ergebnisse über die Konstitution der chromatischen Substanz des Zellkerns [Cell Studies. VI. An introductory treatment of this and other chromosome problems can be found in my article: Findings concerning the constitution of the chromatic substance of the cell nucleus]. *Jena*, (1907).
82. P. C. Nowell, D. A. Hungerford, Chromosome studies on normal and leukemic human leukocytes. *Journal of the National Cancer Institute* **25**, 85-109 (1960); published online EpubJul (
83. J. D. Rowley, Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**, 290-293 (1973); published online EpubJun 1 (
84. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359 (2012); published online EpubApr (10.1038/nmeth.1923).
85. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595 (2010); published online EpubMar 1 (10.1093/bioinformatics/btp698).
86. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010); published online EpubMar 15 (10.1093/bioinformatics/btq033).
87. K. Chen, J. W. Wallis, M. D. McLellan, D. E. Larson, J. M. Kalicki, C. S. Pohl, S. D. McGrath, M. C. Wendl, Q. Zhang, D. P. Locke, X. Shi, R. S. Fulton, T. J. Ley, R. K. Wilson, L. Ding, E. R. Mardis, BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nature methods* **6**, 677-681 (2009); published online EpubSep (10.1038/nmeth.1363).
88. R. M. Layer, C. Chiang, A. R. Quinlan, I. M. Hall, LUMPY: a probabilistic framework for structural variant discovery. *Genome biology* **15**, R84 (2014)10.1186/gb-2014-15-6-r84).
89. B. Zeitouni, V. Boeva, I. Janoueix-Lerosey, S. Loeillet, P. Legoix-ne, A. Nicolas, O. Delattre, E. Barillot, SVDetect: a tool to identify genomic structural variations from paired-end and mate-pair sequencing data. *Bioinformatics* **26**, 1895-1896 (2010); published online EpubAug 1 (10.1093/bioinformatics/btq293).
90. A. L. K. Norris, H; Makohon-Moore, A; Pallavajjala, A; Morsberger, LA; Lee, K; Batista, DA; Iacobuzio-Donahue, CA; Tseh-Lin, M; Klein, AP; Hruban, RH; Wheelan, SJ; Eshleman, JR, TransFlip mutations produce deletions in cancer. *Genes, chromosomes & cancer*, ((submitted)).
91. N. Waddell, M. Pajic, A. M. Patch, D. K. Chang, K. S. Kassahn, P. Bailey, A. L. Johns, D. Miller, K. Nones, K. Quek, M. C. Quinn, A. J. Robertson, M. Z. Fadlullah, T. J. Bruxner, A. N. Christ, I. Harliwong, S. Idrisoglu, S. Manning, C. Nourse, E. Nourbakhsh, S. Wani, P. J.

Wilson, E. Markham, N. Cloonan, M. J. Anderson, J. L. Fink, O. Holmes, S. H. Kazakoff, C. Leonard, F. Newell, B. Poudel, S. Song, D. Taylor, N. Waddell, S. Wood, Q. Xu, J. Wu, M. Pinese, M. J. Cowley, H. C. Lee, M. D. Jones, A. M. Nagrial, J. Humphris, L. A. Chantrill, V. Chin, A. M. Steinmann, A. Mawson, E. S. Humphrey, E. K. Colvin, A. Chou, C. J. Scarlett, A. V. Pinho, M. Giry-Laterriere, I. Rooman, J. S. Samra, J. G. Kench, J. A. Pettitt, N. D. Merrett, C. Toon, K. Epari, N. Q. Nguyen, A. Barbour, N. Zeps, N. B. Jamieson, J. S. Graham, S. P. Niclou, R. Bjerkvig, R. Grutzmann, D. Aust, R. H. Hruban, A. Maitra, C. A. Iacobuzio-Donahue, C. L. Wolfgang, R. A. Morgan, R. T. Lawlor, V. Corbo, C. Bassi, M. Falconi, G. Zamboni, G. Tortora, M. A. Tempero, I. Australian Pancreatic Cancer Genome, A. J. Gill, J. R. Eshleman, C. Pilarsky, A. Scarpa, E. A. Musgrove, J. V. Pearson, A. V. Biankin, S. M. Grimmond, Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495-501 (2015); published online EpubFeb 26 (10.1038/nature14169).

92. A. C. Society., Cancer Facts & Figures 2015. *Atlanta: American Cancer Society;* , (2015).
93. S. Jones, R. H. Hruban, M. Kamiyama, M. Borges, X. Zhang, D. W. Parsons, J. C. Lin, E. Palmisano, K. Brune, E. M. Jaffee, C. A. Iacobuzio-Donahue, A. Maitra, G. Parmigiani, S. E. Kern, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, J. R. Eshleman, M. Goggins, A. P. Klein, Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science* **324**, 217 (2009); published online EpubApr 10 (10.1126/science.1171202).
94. B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, Jr., K. W. Kinzler, Cancer genome landscapes. *Science* **339**, 1546-1558 (2013); published online EpubMar 29 (10.1126/science.1235122).

CURRICULUM VITAE

The Johns Hopkins University School of Medicine

Alexis Leigh Norris March 2015

Educational History

2010-2015 The Johns Hopkins University School of Medicine
Baltimore, MD
Ph.D., Cellular and Molecular Medicine

2004-2007 Millersville University
Millersville, Pennsylvania
B.S. Biology; Biochemistry Minor

Research Experience

Thesis Research (Cancer genomics), Lab of Dr. James R. Eshleman 2011-2015
Johns Hopkins School of Medicine

Research Rotation (DNA damage response), Lab of Dr. Fred Bunz 2011
Johns Hopkins School of Medicine

Research Rotation (Mouse models of cancer), Lab of Dr. Elizabeth Jaffee 2010
Johns Hopkins School of Medicine

Research Specialist (Pancreatic cancer), Lab of Dr. James R. Eshleman 2009-2010
Johns Hopkins School of Medicine

Molecular Technologist, Molecular Diagnostics Lab 2007-2009
Johns Hopkins School of Medicine

Undergraduate Research (Yeast genetics), Lab of Dr. Carol Ely-Hepfer 2006-2007
Millersville University

Academic Honors

Travel Award, Johns Hopkins School of Medicine Graduate Student Association 2015

First Place Poster Award, 8th Annual Young Investigators Symposium on Genomics and Bioinformatics, Johns Hopkins School of Medicine 2014

Neimeyer-Hodgson Research Grant, Millersville University 2007

Excellence in Scholarship, Millersville University 2007

Dean's List, Millersville University 2006-2007

Independent Research Recognition, Millersville University 2006-2007

Publications

Norris AL, Kamiyama H, Makohon-Moore A, Pallavajjala A, Morsberger LA, Lee K, Batista DA, Iacobuzio-Donahue CA, Tseh-Lin M, Klein AP, Hruban RH, Wheelan SJ, Eshleman JR. TransFlip mutations produce deletions in cancer. *Genes Chromosomes Cancer*. (Revision Submitted).

Eshleman JR, **Norris AL**, Sadakari Y, Debeljak M, Borges M, Harrington C, Lin E, Brant A, Barkley T, Almario TJ, Topazian M, Farrell J, Syngal S, Lee JH, Yu J, Hruban RH, Kanda M, Canto MI, Goggins M. KRAS and GNAS mutations in pancreatic juice collected from the duodenum of patients undergoing endoscopic ultrasound for suspected or established pancreatic disease. *Clin Gastroenterol Hepatol*. 2014 Dec 3.

Norris AL, Roberts NJ, Jones S, Wheelan SJ, Papadopoulos N, Vogelstein B, Kinzler K, Hruban RH, Klein AP, Eshleman JR. Familial and sporadic pancreatic cancer share the same molecular pathogenesis. *Fam Cancer*. (In Press).

Chen G, Olson MT, O'Neill A, **Norris A**, Beierl K, Harada S, Debeljak M, Rivera-Roman K, Finley S, Stafford S, Gocke CD, Lin M Eshleman JR. A Virtual Pyrogram Generator to Resolve Complex Pyrosequencing Results. *J Mol Diagn*. 2012 Mar; 14(2):149-59.

Matthaei H*, **Norris AL***, Tsiatis AC, Olino K, Hong SM, dal Molin M, Goggins MG, Canto M, Horton KM, Jackson KD, Capelli P, Zamboni G, Bortesi L, Furukawa T, Egawa S, Ishida M, Ottomo S, Unno M, Motoi F, Wolfgang CL, Edil BE, Cameron JL, Eshleman JR, Schulick RD, Maitra A, Hruban RH. Clinicopathological characteristics and molecular analyses of multifocal intraductal papillary mucinous neoplasms of the pancreas. *Annals Surgery*. 2011 Dec; 255(2):326-33. ***co-first authors.**

Weekes C, Nallapareddy S, Rudek MA, **Norris-Kirby A**, Laheru D, Jimeno A, Donehower RC, Murphy KM, Hidalgo M, Baker SD, Messersmith WA. Thymidylate Synthase Enhancer Region (TSER) Genotype-Directed Phase II Trial of Oral Capecitabine for 2nd Line Treatment of Advanced Pancreatic Cancer. *Invest New Drugs*. 2011 Oct; 29(5):1057-65.

Jinawath N, Morsberger L, **Norris-Kirby A**, Williams LM, Yonescu R, Argani P, Griffin CA, Murphy KM. Complex rearrangement of chromosomes 1, 7, 21, 22 in Ewing sarcoma. *Cancer Genet Cytogenet*. 2010 Aug; 201(1):42-7.

Kamiyama H, Kamiyama M, Hong SM, Karikari C, Lin MT, Griffith M, Young A, **Norris-Kirby A**, Borges M, Mizuma M, Feldmann G, Shi C, Liang H, Goggins M, Maitra A, Hruban RH, Eshleman JR. In vivo and in vitro propagation of intraductal papillary mucinous neoplasms. *Lab Invest*. 2010 May; 90(5):665-73.

Tsiatis AC, **Norris-Kirby A**, Rich RG, Hafez MJ, Gocke CD, Eshleman JR, Murphy KM. Comparison of Sanger Sequencing, Pyrosequencing, and Melting Curve analysis for the detection of KRAS Mutations: Diagnostic and Clinical Implications. *J Mol Diagn*. 2010 Jul; 12(4)425-32.

Norris-Kirby A, Hagenkord JM, Kshirsagar MP, Ronnett BM, Murphy KM. Abnormal villous morphology associated with triple trisomy of paternal origin. *J Mol Diagn*. 2010 Jul; 12(4):525-9.

Jinawath N, **Norris-Kirby A**, Smith BD, Gocke CD, Batista DA, Griffin CA, Murphy KM. A rare e14a3 (b3a3) BCR-ABL fusion transcript in chronic myeloid leukemia: diagnostic challenges in clinical laboratory practice. *J Mol Diagn*. 2009. 11(4):359-63.

McConnell TG, **Norris-Kirby A**, Hagenkord JM, Ronnett BM, Murphy KM. Complete hydatidiform mole with retained maternal chromosomes 6 and 11. *Am J Surg Path*. 2009. 33(9):1409-15.

Abstracts and Presentations

Invited speaker and instructor. Hagerstown Community College (Maryland). January 16-17, 2015.

Norris AL and Eshleman JR. Tumor suppressor genes are inactivated by complex rearrangements, 8th Annual Young Investigators Symposium on Genomics and Bioinformatics, October 2014 in Baltimore, MD (poster presentation and first place award).

Norris AL, Roberts NR, Hruban RH, Klein AP, Eshleman JR. Shared driver genes of familial and sporadic pancreatic cancer may explain the similar age of onset, 64th Annual Meeting of the American Society of Human Genetics, October 2014 in San Diego, CA (poster presentation).

Invited speaker and representative of The Sol Goldman Pancreatic Cancer Research Center at Johns Hopkins. Purplepalooza Invitational. Cattail Country Club, Glencoe, MD. September 2013 (1st annual) and September 2014 (2nd annual).

Invited instructor. Fallston High School (MD). April 2014.

Norris AL, Roberts NR, Hruban RH, Klein AP, Eshleman JR. Shared driver genes of familial and sporadic pancreatic cancer may explain the similar age of onset, 64th Annual Meeting of the American Society of Human Genetics, October 2014 in San Diego, CA (poster presentation).

Chen G, Olson M, **Norris-Kirby A**, Beierl K, Harada S, Debeljak M, Rivera-Roman K, Finley S, Stafford A, Gocke CD, Lin MT, Eshleman JR. A virtual pyrogram generator to resolve ambiguity of pyrosequencing results, 103rd Annual Meeting of the American Association for Cancer Research, April 2012 in Chicago, IL.

Chen G, Olson M, **Norris-Kirby A**, Harada S, Debeljak M, Rivera-Roman K, Beierl K, Stafford A, Gocke CD, Lin MT, Eshleman JR. Integrated analysis of two technologies resolves ambiguity of DNA sequencing results, 101st Annual Meeting of the United States and Canadian Academy of Pathology, March 2012 in Vancouver, BC.

Chen G, Olson M, **Norris-Kirby A**, Harada S, Debeljak M, Rivera-Roman K, Beierl K, Stafford A, Gocke CD, Lin MT, Eshleman JR. Integrated analysis of two technologies resolves ambiguity of DNA sequencing results, 17th Annual Meeting of the Association for Molecular Pathology 2009 Annual Meeting, November 2011 in Grapevine, TX.

Jinawath N, Murphy K, **Norris A**, Klein A, Yonescu R, Iacobuzio-Donahue C, Brody J, Meeker A, Jinawath A, Harada S, Griffin C. CDH10 mutation profiles in pancreatic ductal adenocarcinomas (PDAC), 60th Annual Meeting of the American Society of Human Genetics, November 2010 in Washington, D.C.

Matthaei H, **Norris-Kirby A**, Tsiatis AC, Hong SM, Olino K, Wolfgang CL, Schulick RD, Cameron JL, Jackson KD, Capelli P, Zamboni G, Bortesi L, Furukawa T, dal Molin M, Goggins M, Eshleman JR, Hruban RH, A. Maitra, Multifocal IPMNs: pathology and field defect, 41st Annual Meeting of the American Pancreatic Association, November 2010 in Chicago, IL.

Tsiatis AC, **Norris-Kirby A**, Rich RG, Hafez MJ, Gocke CD, Eshleman JR, and Murphy KM. Comparison of Sanger Sequencing, Pyrosequencing, and Melting Curve analysis for the detection of KRAS Mutations: Diagnostic and Clinical Implications, 15th Annual Meeting of the Association for Molecular Pathology 2009 Annual Meeting, November 2009 in Kissimmee, FL.

Norris-Kirby A, Hagenkord JM, Kshirsagar MP, Ronnett BM, and Murphy KM. Abnormal villous morphology associated with triple trisomy of paternal origin. 15th Annual Meeting of the Association for Molecular Pathology 2009 Annual Meeting, November 2009 in Kissimmee, FL (poster presentation).

Norris-Kirby A, McConnell TG, Hagenkord JM, Ronnett BM, and Murphy KM. Complete hydatidiform mole with retained maternal chromosomes 6 and 11. 18th Annual Meeting of the American College of Medical Genetics, March 2009 in Tampa, FL (poster presentation).

Service and Leadership Experience

STEM Specialist, Maryland Business Roundtable for Education	2014-2015
PanCAN National Capital Affiliate Volunteer Chair	2009-2015
Incentive Mentoring Program (IMP) Volunteer Chair	2010-2011